

2006 Halvor Schrøder Hansen

Master's thesis

Modelling of carbohydrates with a reactive force field

NTNU  
Norwegian University of  
Science and Technology  
Faculty of Natural Sciences and Technology  
Department of Chemistry

Halvor Schrøder Hansen

## Modelling of carbohydrates with a reactive force field

Masteroppgave til graden  
Master i teknologi/sivilingeniør

Spring 2006

Supervisor: Professor Per-Olof Åstrand



## Statement of independent work

I hereby declare that this thesis has been carried out independently and in agreement with "Forskrift om krav til mastergrad" established by the Norwegian Ministry of Education and Research, 01.12.2005, and in agreement with supplementary regulations concerning Master theses at the Norwegian University of Science and Technology (NTNU).

Date:

---

Halvor Schrøder Hansen

# Modelling of carbohydrates with a reactive force field

Halvor Schrøder Hansen

June 5, 2006

## Abstract

A primer to a reactive force field based on the ReaxFF formulation has been developed by expanding the ReaxFF hydrocarbon force field to include oxygen. Energies of relevant carbohydrate structures and reaction paths, together with model systems for oxidation and protonation have been calculated by DFT methods together with a few Mulliken charges and some important bond lengths and angles. The structures and the corresponding energies, charges and geometrical parameters have been joined with a large number of structures and reactions involving small carbon-hydrogen-oxygen molecules and used for parametrisation of the oxygen parameters of the ReaxFF force field. The force field is tested and discrepancies in the parametrisation have been identified and described. The primer is able to describe the reaction barriers and predicts reasonable stabilities for most of the systems investigated, but fails on two main points, firstly it overstabilises carbocations and secondly it does not predict the correct effect that  $\text{H}_2\text{O}$  has on the stability of  $\text{H}_3\text{O}^+$  and on the energy barriers of protonation. It is believed that these discrepancies can be removed by a reparametrisation of the force field by including carbocation structures with corresponding DFT-energies and protonation reaction barriers calculated with non-reacting water molecules present into the training set, and also by refining a set of  $\text{H}_2\text{O}/\text{H}_3\text{O}^+$  clusters already included in the first training set.

Three new algorithms have been implemented in the ReaxFF program package, a Root Mean Squared Distance fit, a Quasi-Newton method for energy minimisation based on the BFGS update and a Nudged Elastic Band method for the calculation of reaction paths. A new formulation of the proof of how to remove forces that lead to translation and rotation of a molecule is also presented, together with an extended proof of the Root Mean Squared Distance algorithm based on a Singular Value Decomposition. The theoretical background as well as examples of the implemented methods are provided.

## Preface

This project is the final project to conclude a Master of Science/”Master i Teknologi” degree in chemistry at the Norwegian University of Science and Technology (NTNU), Trondheim. It has been performed during a period of 20 weeks and covers 30 ECTS points. This project has been performed at the Department of Chemistry, NTNU, with Professor Per-Olof Åstrand as supervisor. Also included in this project has been a two-week visit to the California Institute of Technology (Caltech), at the Materials and Process Simulation Center, directed by Professor William A. Goddard, III. My supervisor at Caltech has been Adri van Duin, Director of Force Fields and Atomistic Simulations at the Materials and Process Simulation Center.

The project was motivated by a request from SINTEF Energy Research, where they study the aging of cellulose in transformers experimentally[1]. Cellulose is commonly used as an insulator in transformers, and over time, the cellulose chains degrade and their insulation properties are reduced. To get more insight into the aging process, they were interested in using atomistic simulations as a tool to investigate possible reactions and ultimately how to inhibit the degradation.

However, before such simulations can be performed, we need to develop a tool to describe the reactions that can occur in cellulose. The calculation of energies of atoms and molecules can be done by the use of quantum chemistry methods, but in simulations we have to calculate the energy for every time step and we are also interested in describing large systems. A typical simulation could contain about 10000 atoms and include millions of time steps. Quantum chemistry methods are computationally demanding, and limit us to systems of the size of about 50 atoms for only a few steps. The chosen approach for simulations is then to use a force field, which approximates quantum chemistry energies by analytical functions. These analytical functions are much faster to solve than solving the quantum chemistry problem, and we are able to run larger system and more time steps.

There is a number of force fields developed for atomistic simulations (e.g. in GROMOS[2], AMBER[3], GROMACS[4] and NAMD[5]<sup>1</sup>), but most force fields are directed towards describing structure and conformations of molecules, and do not allow any reactions to occur. Since we are interested in describing the reactions that occur in cellulose, we need to develop a force field that includes reactions. Since cellulose itself is a carbohydrate, the overall target of this project is to develop a reactive force field for carbohydrates.

The development of force fields is an iterative process. The analytical terms in the force fields need to be parametrised to some data, this could be quantum chemistry energies, experimental bond lengths, heats of formation etc. There is an enormous amount of possible events that can occur in a system of atoms and it is impossible to model all of these by quantum chemistry, thus we have to choose a limited set of data to parametrise our force field to, and hope that we by that can catch the overall trends and that the force field will be valid for similar systems. There will always be some effects that are not described properly by the force field, as we approximate a very complicated system by a limited set of analytical functions. After a parametrisation has been performed, the force field has to be tested to see if it gives reasonable results. If not, the discrepancies, if possible, should be removed by a new parametrisation, by fitting to an improved data set. The aim of this project is then to develop a primer to such a force field, and identify discrepancies and errors in the primer, thus identifying which additional data that should be included in the parametrisation and to point out which effects the force field is able to describe.

The chosen reactive force field is a force field called ReaxFF developed for hydrocarbons, which has been developed at the Materials and Process Simulation Center, Caltech[6]. ReaxFF applies a set of analytical

---

<sup>1</sup>All these articles are in fact published in the same issue of J. Comput. Chem. (volume 26, issue 16), 2005 and is recommended to read for a reference on the most used tools for MD-simulations

energy functions, which has to be parametrised to include carbohydrates. ReaxFF also comes with a program package which includes methods for energy minimisation, force field optimisation and molecular dynamics (MD) simulations. However, it is also desirable to expand this package by including other useful methods. A substantial part of this project has therefore been employed on method development, and several features have been added to the ReaxFF program package.

The methods implemented in the ReaxFF program package are well known methods and have been employed for other applications. However, during the implementation a couple of questions to the mathematical background of some of the methods arose. An extensive version of a proof of the Root Mean Squared Distance (RMSD) fit procedure given by Ho[7] has been formulated, which is believed to be an easier understandable proof than the original proof by Kabsch[8]. Also a new proof on how to remove forces that would lead to overall translation and rotation of a molecule has been formulated for use in the implemented Nudged Elastic Band (NEB)[9, 10] method. The programming details of the implemented methods are not included in this report, but examples of the RMSD-fit and the NEB-method are included.

As a summary, this project deals with

1. 1<sup>st</sup> parametrisation of force field  
Energies, structures and Mulliken[11] charges for relevant carbohydrate systems and reactions have been calculated by the use of quantum chemistry. These energies and structures have been included in a training set for ReaxFF for the expansion of the present hydrocarbon force field to a carbohydrate force field. The parametrisation itself has been performed by Adri van Duin at the Materials and Process Simulation Center, Caltech.
2. Testing of force field  
The 1<sup>st</sup> parametrised force field has been tested and defects and discrepancies of the 1<sup>st</sup> parametrisation have been identified and suggestions of supplementary data to include in the training set for the next parametrisation have been given.
3. Development of methods  
Implementation of a RMSD-fit algorithm, a Quasi-Newton method for energy minimisation and a Nudged Elastic Band method for calculation of reaction paths. The implemented methods have been described and examples of the use of these methods have been given. An extended version of the proof of the RMSD-fit procedure given by Ho[7], has been formulated, together with a new formulation of the proof of how to remove forces that would lead to overall translation and rotation of a molecule.

## Acknowledgements

I would especially like to thank Adri van Duin at the Materials and Process Simulation Center, Beckman Institute, Caltech, California for help and supervision during this master project. I also appreciate his company during my stay at Caltech in the two first weeks of March which I enjoyed a lot. I would also thank William A. Goddard III for accepting me as a visitor to the Materials and Process Simulation Center, Caltech.

I also thank Frank A. Momany at Plant Polymer Research, ARS, National Center for Agricultural Utilization Research, Illinois, USA for providing coordinates for glucose conformations. I would also like to thank my supervisor Per-Olof Åstrand for providing the task and helpful suggestions concerning this report. Finally I would like to thank Terje Tofteberg for last minute comments and Cathryn E. Birch for correcting my english.

I am also grateful to Sandra Gilles for sending me "spätzli" and chocolate all the way from Germany, and to my students in the Basic Physical Chemistry class who provided me with a bag of "Twist" (also chocolate). This has helped me a lot the final weekend, and enabled me to work long days and finish this project.

# Contents

<b>Preface</b>	<b>i</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Basic Theory</b>	<b>2</b>
2.1 Introduction to Quantum Mechanics . . . . .	2
2.2 Approximations . . . . .	4
2.3 Electronic Structure Methods . . . . .	5
2.4 Basis sets . . . . .	13
<b>3 Force Fields</b>	<b>15</b>
3.1 Common force field terms . . . . .	16
3.2 Reactive Force Fields . . . . .	17
3.3 ReaxFF Force Field Terms . . . . .	18
3.4 Electronegativity Equalisation . . . . .	22
3.5 Simulations . . . . .	23
<b>4 Algorithms</b>	<b>24</b>
4.1 RMSD fit . . . . .	24
4.2 Removal of translation and rotation . . . . .	27
4.3 Quasi-Newton method . . . . .	29
4.4 Nudged elastic band . . . . .	30
<b>5 Computational details</b>	<b>32</b>
5.1 Quantum chemistry calculations . . . . .	32
5.2 Parametrisation . . . . .	32
5.3 Test of force field . . . . .	33
<b>6 Results and discussion</b>	<b>33</b>
6.1 Parametrisation and comparison to DFT-data . . . . .	33
6.2 Test of force field . . . . .	42
6.3 Where to go next? . . . . .	47
<b>7 Conclusion</b>	<b>48</b>
<b>Appendix</b>	<b>53</b>
<b>A ReaxFF Energy terms</b>	<b>53</b>
A.1 Valence angle energy . . . . .	53
A.2 Torsion angle energy . . . . .	54
A.3 Nonbonded interactions . . . . .	54
A.4 Special terms . . . . .	55
A.5 ReaxFF parameters . . . . .	56

# 1 Introduction

Carbohydrates are one of the most abundant types of organic compounds we have. Most biomass is made up of carbohydrates, more than 50 % of the dry weight of plants[12], and they are commonly used as structural units (e.g. cellulose for plants and chitin for insects) and as energy storage (e.g. starch and sugars). Carbohydrates are important everyday substances, as they are a source of 80 % of the total food intake for the worlds population, and they are important substances in our body, e.g. as structural units of DNA and antigens that are recognised by our immuno defence system.

Carbohydrates have a wide range of uses, ranging from the paper industry to materials, biopolymers, food additives, fuel, beverages etc. It has also recently been identified as the next "frontier" in biomedicine [13], being the area where medical advances are expected in the coming years, e.g. carbohydrates are believed to be important in the search for a HIV vaccine[14].

Among the many applications of carbohydrates, oil-impregnated paper has been used as an insulator in power transformers. However, over time the cellulose chains depolymerise and their insulating ability decreases. This can then lead to short circuits and destruction of the transformer. To increase the life time of the transformer, we would like to inhibit this destruction process. Experiments have been performed to study the aging of cellulose[1], but to understand more about what is happening in the aging process, molecular simulations can be a useful tool.

Molecular simulations can in principle be performed by using quantum chemistry methods, but these methods are computationally infeasible for all but small systems for short periods of time. Quantum chemistry methods are commonly employed on the calculation of simple reaction paths such as the ring opening of glucose[15]. To make it possible to perform simulations, we approximate the energy calculated by quantum chemistry by using a set of analytical energy terms. This set of analytical energy terms is called a "force field". There exist many different kinds of force fields for many different applications, and the most common force fields are of the non-reactive type, thus atoms stay bonded all the time. These are used to study conformational changes, transport processes etc., and have been commonly applied to carbohydrates (e.g. [16] and [17]). One approach to include reactions is to treat a small part of our system with quantum mechanics and the rest by a force field, and these hybrid methods are denoted QM/MM(Quantum Mechanics/Molecular Mechanics) and one of these methods has also been applied on the degradation of glucose[18]. However, this is still computationally demanding and it is desirable to also treat reactions with analytical energy terms. The development of reactive force fields has been proceeding in parallel with the non-reactive force fields the last two decades with the introduction of bond-order potentials by Abell in 1985[19] which were further expanded by Tersoff[20] in 1988 and Brenner[21] in 1990. These types of potentials are usually called "reactive empirical bond-order" (REBO) potentials. Lately what has been denoted the 2<sup>nd</sup> generation reactive force fields have been developed e.g. AIREBO by Stuart et al.[22], REBO II by Brenner et al.[23] and ReaxFF by van Duin et al.[6]. These include modified and improved potential functions and have been applied on numerous problems. ReaxFF has for instance been employed on problems ranging from hydrocarbons[6] to silicon[24], transition metal catalysis[25], explosives[26, 27, 28] and even on such a thing as a maille made of nanotubes[29].

Carbohydrates display a very complicated Potential Energy Surface (PES). They contain a large number of hydroxy groups which are able to make hydrogen bonds to each other, they can exist in ring structures, open structures and they can form a large variety of shapes. There are also a large number of reactions that can occur in carbohydrates, e.g. for the pyrolysis of cellulose, 96 different products have been identified[30]. Hopefully, a parametrisation of the bond-order potential as defined by ReaxFF[6] is able to approximate this PES at an accuracy good enough to give the right trends in predicting possible reactions involving carbohydrates, thus making it an useful tool for carbohydrate research.

## 2 Basic Theory

To be able to perform simulations of particles such as atoms, we need to describe how nature works. Nature itself tends to lower its energy, an observation made by Isaac Newton thus proposing the famous relation between forces and movement of items, also called Newton's 2nd law.

$$F = ma \quad (2.1)$$

or the more mathematical

$$-\frac{dV}{dr} = m \frac{d^2r}{dt^2} \quad (2.2)$$

where  $V$  is the potential in where the item is moving. This basically enables us to predict movements if we are able to describe the energy of our system.

Newton's equation (equation 2.1), gives a correct description for most everyday problems, but for items moving close to the speed of light and small scale particles, modified sets of equations have to be used. For items close to the speed of light, we use a relativistic correction to equation 2.1 proposed by Einstein which is not necessary in this project. Our interest lies in particles of the size of atoms, nuclei and electrons and these have a certain character of a wave, and for such cases equation 2.1 was reformulated to a wave problem, given by the Schrödinger equation

$$\hat{H}\psi = i\hbar \frac{\partial\psi}{\partial t} \quad (2.3)$$

where  $\hat{H}$  is the Hamilton operator, an operator which gives the energy of the system and will be described further in Chapter 2.1. This assumption of particles being described as waves, is denoted quantum mechanics as opposed to classical mechanics, where the Newton equation 2.1 is being solved. We are interested in describing atoms, containing nuclei and electrons, and such small particles must be described by using the wave assumption, therefore quantum mechanics has to be employed.

### 2.1 Introduction to Quantum Mechanics

There are many books concerning quantum mechanics. For a beginner's introduction the book of McQuarrie and Simon[31] is recommended. An overview of common computational methods is given in the book of Jensen[32]. Density functional theory is well described in the article by Capelle[33] and by Kohanoff and Gidopoulos[34]. This introduction to quantum mechanics is based on these sources.

Quantum mechanics is based on the assumption that particles behave like waves. There are 5 fundamental postulates that create the basis of the quantum mechanics which are listed below.

1. The quantum mechanical system is completely described by its wave function  $\psi$ . No more information is required to calculate the properties of the system.
2. To every observable in classical mechanics there is a linear operator ( $\hat{A}$ ).
3. The observable will only obtain values that corresponds to Eigenvalues ( $a_n$ ) and the corresponding Eigenstates ( $\psi_n$ ) of the these operators ( $\hat{A}$ )

$$\hat{A}\psi_n = a_n\psi_n \quad (2.4)$$

4. The average value of the observable is given by

$$\bar{A} = \int_{\text{all space}} \psi^* \hat{A} \psi dx = \langle \psi | \hat{A} | \psi \rangle \quad (2.5)$$

5. The wave function evolves in time according to

$$\hat{H} \psi = i\hbar \frac{\partial \psi}{\partial t} \quad (2.6)$$

which is denoted the time-dependent Schrödinger equation, and the  $\hat{H}$  is called the Hamilton operator.

With the use of these postulates, we are in principle able to calculate the energy (and many other properties) and the time evolution of the system exactly. For our purpose, we are mostly interested in the energy due to a specific configuration of electrons and nuclei. That means we have to obtain a wave function, and apply the energy operator  $\hat{H}$  (the Hamilton operator) to our system.

### 2.1.1 The Hamilton operator

The Hamilton operator is an operator that gives the energy of our wave function. It consists of two parts, the kinetic energy  $\hat{T}$  and the potential energy  $\hat{V}$

$$\hat{H} = \hat{T} + \hat{V} \quad (2.7)$$

Our potential energy arises from the charge-charge interactions between electrons and nuclei

$$\hat{V} = \sum_{i < j}^{N_e} \frac{e^2}{4\pi\epsilon_0 |r_i - r_j|} - \sum_{ij}^{N_e N_n} \frac{e^2 Z(j)}{4\pi\epsilon_0 |r_i - r_j|} + \sum_{i < j}^{N_n} \frac{e^2 Z(i) Z(j)}{4\pi\epsilon_0 |r_i - r_j|} \quad (2.8)$$

where  $N_e$  and  $N_n$  are the respective numbers of electrons and nuclei,  $-e$  is the electron charge and the charge of a nuclei is given as  $Z(i)e$ . The kinetic energy operator is given as

$$\hat{T} = - \sum_i^{N_n} \frac{\hbar}{2m_{n_i}} \nabla_{n_i}^2 - \sum_j^{N_e} \frac{\hbar}{2m_e} \nabla_{e_j}^2 \quad (2.9)$$

where the first term describes the kinetic energy of the nuclei and the last term the kinetic energy of the electrons.  $m_x$  describes here the mass of the electron/nuclei and the  $\nabla^2$  operator is the familiar differentiation operator ( $\nabla^2 = (\frac{\partial}{\partial x^2}, \frac{\partial}{\partial y^2}, \frac{\partial}{\partial z^2})$ ). The coordinates of the electrons and the nuclei are all allowed to vary over all space, such that our wave function will be dependent on a large set of coordinates and also time, as

$$\psi = \psi(r_1^e, r_2^e, \dots, r_{N_e}^e, r_1^n, r_1^n, \dots, r_{N_n}^n, t) = \psi(\mathbf{r}^e, \mathbf{r}^n, t) \quad (2.10)$$

where all the electron coordinates are gathered into  $\mathbf{r}^e$  and all nuclei coordinates into  $\mathbf{r}^n$ .

### 2.1.2 The time-independent Schrödinger equation

When our Hamilton operator does not explicitly contain time, we can reformulate equation 2.6 to an equation that does not contain time by assuming that the time-part of our wave function can be factored out such that

$$\psi(\mathbf{r}^e, \mathbf{r}^n, t) = \psi(\mathbf{r}^e, \mathbf{r}^n) f(t) \quad (2.11)$$

and that our Hamilton operator does not depend on time such that

$$\hat{H}\psi(\mathbf{r}^e, \mathbf{r}^n)f(t) = f(t)\hat{H}\psi(\mathbf{r}^e, \mathbf{r}^n) \quad (2.12)$$

By using equation 2.6 we obtain

$$\frac{1}{\psi(\mathbf{r}^e, \mathbf{r}^n)}\hat{H}\psi(\mathbf{r}^e, \mathbf{r}^n) = \frac{i\hbar}{f(t)}\frac{df(t)}{dt} = \lambda(\text{constant}) \quad (2.13)$$

The  $\lambda$  has to be constant since the left side of equation 2.13 depends on  $\mathbf{r}^e$  and  $\mathbf{r}^n$  and the right side depends on  $t$ . The left-hand side of this equation separates to

$$\hat{H}\psi(\mathbf{r}^e, \mathbf{r}^n) = \lambda\psi(\mathbf{r}^e, \mathbf{r}^n) = E\psi(\mathbf{r}^e, \mathbf{r}^n) \quad (2.14)$$

which is called the time-independent Schrödinger equation. Note that the constant  $\lambda$  has been changed to  $E$  which equals the energy, as the Hamilton operator is the energy operator. Note also that this is an Eigenvalue problem of the same type as equation 2.4. This equation has many solutions, each one corresponding to different Eigenfunctions and Eigenvalues and one therefore usually write equation 2.14 with indices as

$$\hat{H}\psi_i(\mathbf{r}^e, \mathbf{r}^n) = E_i\psi_i(\mathbf{r}^e, \mathbf{r}^n) \quad (2.15)$$

and these corresponds to different excited states. We are interested in describing the ground state, thus the state with the lowest energy  $E_0$ .

## 2.2 Approximations

Equation 2.14 is not trivial to solve and does in general not give analytical solutions. The wave function is dependent on the configurations of both nuclei and electrons which in principle are all waves in three dimensions. We will therefore apply some approximations to be able to solve this equation and obtain the energy of our system.

### 2.2.1 The Born-Oppenheimer approximation

As the electrons have a much smaller mass than the nuclei (about  $\frac{1}{2000}$  the mass of a proton), their velocity will be much higher than the velocity of the nuclei. Or in other words, if we were treating the particles classically, equation 2.2 would give  $> 2000$  times higher acceleration for the electrons than for the nuclei. This enables us to treat the movements of nuclei and electrons independently, by assuming that the nuclei are motionless compared to the electrons. Mathematically this can be written as

$$\psi(\mathbf{r}^e, \mathbf{r}^n) = \psi_e(\mathbf{r}^e; \mathbf{r}^n)\psi_n(\mathbf{r}^n) \quad (2.16)$$

where now  $\psi_e$  is the wave function for the electrons, which depends on the nuclear coordinates as parameters. In other words, the electronic wave function can be solved by keeping the nuclear coordinates fixed. This approximation is called the Born-Oppenheimer approximation and is a very good approximation and leads to negligible errors for most cases. We are then able to split the Schrödinger equation (equation 2.14) into an electronic part and a nucleic part, namely

$$\hat{H}\psi = \hat{H}_e\psi_e(\mathbf{r}^e; \mathbf{r}^n)\psi_n(\mathbf{r}^n) + \hat{H}_n\psi_e(\mathbf{r}^e; \mathbf{r}^n)\psi_n(\mathbf{r}^n) = (V_e + V_n)\psi_e(\mathbf{r}^e; \mathbf{r}^n)\psi_n(\mathbf{r}^n) \quad (2.17)$$

where

$$\hat{H}_e = -\sum_j^{N_e} \frac{\hbar}{2m_e} \nabla_{e_j}^2 + \sum_{i<j}^{N_e} \frac{e^2}{4\pi\epsilon_0|r_i - r_j|} - \sum_{ij}^{N_e N_n} \frac{e^2 Z(j)}{4\pi\epsilon_0|r_i - r_j|} \quad (2.18)$$

and

$$\hat{H}_n = - \sum_i^{N_n} \frac{\hbar}{2m_{n_i}} \nabla_{n_i}^2 + \sum_{i<j}^{N_n} \frac{e^2 Z(i)Z(j)}{4\pi\epsilon_0 |r_i - r_j|}. \quad (2.19)$$

and  $V_e$  and  $V_n$  are the Eigenvalues of the two parts, thus corresponding to the electronic and the nucleic energies. To obtain the electronic energy we need to solve the first part of equation 2.17,

$$\hat{H}_e \psi_e(\mathbf{r}^e; \mathbf{r}^n) \psi_n(\mathbf{r}^n) = \psi_n(\mathbf{r}^n) \hat{H}_e \psi_e(\mathbf{r}^e; \mathbf{r}^n) = V_{el}(\mathbf{r}^n) \psi_e(\mathbf{r}^e; \mathbf{r}^n) \psi_n(\mathbf{r}^n) \quad (2.20)$$

which simplifies to

$$\hat{H}_e \psi_e(\mathbf{r}^e; \mathbf{r}^n) = V_{el}(\mathbf{r}^n) \psi_e(\mathbf{r}^e; \mathbf{r}^n). \quad (2.21)$$

As the nuclei are (still) much heavier than the electrons, we choose to avoid the use of quantum mechanics on the nuclei, and return to classical mechanics. This gives us the following energy  $V_n$  for the nuclei

$$V_n = \sum_i \frac{m_{n_i} v_i^2}{2} + \sum_{i<j} \frac{e^2 Z(i)Z(j)}{4\pi\epsilon_0 |r_i - r_j|} \quad (2.22)$$

and the total energy of our system becomes

$$V_{tot} = \sum_i \frac{m_{n_i} v_i^2}{2} + \sum_{i<j} \frac{e^2 Z(i)Z(j)}{4\pi\epsilon_0 |r_i - r_j|} + V_{el}(\mathbf{r}^n) \quad (2.23)$$

The Born-Oppenheimer approximation is for most cases good, but some questions arise from the lighter nuclei, especially hydrogen, thus leading to some errors in the description of dense phases of hydrogen and hydrogen-bonded systems, but is employed here without further discussion. Equation 2.23 gives us the energy as a function of the nuclear coordinates. If we are able to describe this energy, we can also let the nuclei propagate classically due to Newton's 2nd law (equation 2.2), and thereby predict the time evolution of our system. The main problem in equation 2.23 is to obtain the electronic energy  $V_{el}(\mathbf{r}^n)$ , where we have to use quantum mechanics, by finding the ground state of equation 2.21. The calculation of this electronic energy is the main focus of electronic structure methods such as, "perturbation theory", "configuration interaction" and "density functional theory" (DFT). The science of searching for atomic and molecular properties like energies, geometries, vibrational modes etc. by using quantum mechanics is denoted quantum chemistry.

## 2.3 Electronic Structure Methods

Our use of quantum mechanics is limited to the determination of the electronic wave function given a configuration of nuclei which create a potential for the electrons to move in. By solving the electronic Schrödinger equation (equation 2.21) this potential is assumed constant by the Born-Oppenheimer approximation. We therefore treat this potential as an external potential. By renaming the terms in equation 2.18 and combining with equation 2.21 we get

$$\hat{H}_e \psi_e(\mathbf{r}^e; \mathbf{r}^n) = (\hat{T} + \hat{V}_{ee} + \hat{V}_{ext}) \psi_e(\mathbf{r}^e; \mathbf{r}^n) = V_{el} \psi_e(\mathbf{r}^e; \mathbf{r}^n) \quad (2.24)$$

where  $\hat{V}_{ee}$  is the potential between electrons, and  $\hat{V}_{ext}$  is the external potential which equals the potential created by the nuclei when there is no real external potential. The total energy itself can be found by integrating over all space due to equation 2.5 such that

$$V_{el} = \langle \psi_e | \hat{H}_e | \psi_e \rangle = \langle \psi_e | \hat{T} | \psi_e \rangle + \langle \psi_e | \hat{V}_{ee} | \psi_e \rangle + \langle \psi_e | \hat{V}_{ext} | \psi_e \rangle \quad (2.25)$$

This integration will also return the average energy if our wave function is a superposition of the Eigenfunctions of equation 2.15. If our wave function really is in its ground state, the integration procedure will

only return the Eigenvalue corresponding to the ground state wave function. We will seek to minimise the average energy as defined in equation 2.25 to attempt to find the ground state energy, and we do this by approximating the wave function and iterating to better solutions. Many methods to approximate the wave function have been used, and are well studied and employed on numerous problems. A description of many such methods can be found in Jensen[32]. This report will only handle the method of Density Functional Theory, as this is the method employed in this project.

### 2.3.1 Density Functional Theory

One observable we can obtain from the wave equation for the electrons (equation 2.20) is the electron density. For a one-electron wave function, this is defined as

$$\rho(r) = \psi(r)\psi^*(r) \quad (2.26)$$

where the \* denotes the complex conjugate. However, for a system with more than one electron, we need to integrate over all but one coordinate such that the density becomes

$$\rho(r) = \int \dots \int \psi(r, r_2, r_3, \dots, r_{N_e})\psi^*(r, r_2, r_3, \dots, r_{N_e})dr_2dr_3\dots dr_{N_e} \quad (2.27)$$

By conventional electronic structure methods, the density can be calculated once the wave function  $\psi$  has been obtained. The main idea behind density functional theory is that the electron density does contain as much information as the wave function itself. This is reasonable, since it has been proved that for a given ground state electron density there exists a corresponding external potential. By knowing the external potential, we can get the wave functions by solving equation 2.24. So since we can obtain the nuclear coordinates from the density, we are also able to obtain the wave function. A conventional wave function method assumes that

$$V_{ext} \rightarrow \psi \rightarrow \text{observables} \quad (2.28)$$

thus a wave function is uniquely defined by the external potential, which again can be used to get observables. DFT methods approach the problem as

$$\rho \rightarrow V_{ext} \rightarrow \psi \rightarrow \text{observables} \quad (2.29)$$

thus, the potential is uniquely defined for a ground state density, and by equation 2.28, so is also the wave function and all its observables. It is also to be noted that not only the ground state wave function is uniquely defined by the ground state density, but all the Eigenfunctions of equation 2.15, since they are also a result of the external potential.

The proof of the existence of the one-to-one relationship between the ground state electron density and the external potential (and also the ground state wave function) was formulated by Hohenberg and Kohn and can be found in their original paper[35] or in most DFT books (e.g. [34]). A valuable example for how the external potential can be found from the particle density was presented by Kato[36] and states that the density has a cusp at the position of the nuclei. By finding the position of this cusp, we find the position of the nuclei. In addition, the charge of each nuclei can be found from

$$Z_k = -\frac{a_0}{2\rho(r)} \frac{d\rho(r)}{dr} \Big|_{r=r_n} \quad (2.30)$$

where  $a_0$  equals the bohr radius  $\frac{4\pi\epsilon_0\hbar^2}{m_e e^2}$ . In a coulombic potential, the charges and position of the nuclei determine the external potential completely. In other words, if we know the ground state electron density, we know the position of nuclei, and can calculate the wave function and all it's properties. This demonstrates

that the ground state electron density contains as much information as the wave function.

If we look at equation 2.25 we see that the electric energy is given if we know the wave function, thus it is a functional of the wave function

$$V_{el} = V_{el}[\psi_e] = T[\psi_e] + V_{ee}[\psi_e] + V_{ext}[\psi_e] \quad (2.31)$$

and the dependence is given in equation 2.25. Since the wave function is uniquely defined by the electron density, the electronic energy is also given as a functional of the electron density as

$$V_{el}[\rho(r)] = T[\rho(r)] + V_{ee}[\rho(r)] + V_{ext}[\rho(r)] \quad (2.32)$$

The advantage of this formulation is that the electron density only depend on one coordinate ( $r$ ), while the wave function depends on a set of coordinates ( $r_1, r_2 \dots r_{N_e}$ ), one for each electron. Thus we can attempt to find the particle density instead of the wave function on our way to obtain the electronic energy. However, to determine whether this formulation is advantageous or not, we need to have a way to find the particle density, and we also need to know what the functionals in equation 2.32 look like.

### 2.3.2 Note on spin

Density functional theory often splits the electron density in two parts of different spin

$$\rho(r) = \rho_{\uparrow}(r) + \rho_{\downarrow}(r) \quad (2.33)$$

However, this generalization has not been considered here, as it would complicate the equations without bringing extra understanding. Thus all equations in this paper are in terms of the total electron density unless specified otherwise.

### 2.3.3 Variational principle

Equation 2.15 gives a complete set of orthogonal wavefunctions and any trial wavefunction can be written as a linear combination of those and we get

$$\psi_e^{trial} = \sum_i^{\infty} c_i \psi_e^i \quad (2.34)$$

where the  $\psi_e^i$ 's are the Eigenfunctions of equation 2.15. The energy of the trial wave function will then be (from equation 2.5)

$$V_{el}[\psi_e^{trial}] = \langle \psi_{trial} | \hat{H}_e | \psi_{trial} \rangle = \sum_i |c_i|^2 V_{el}^i[\psi_e^i] \quad (2.35)$$

To maintain the normalisation we need that

$$\sum_i |c_i|^2 = 1 \quad (2.36)$$

and if we assume that  $V_{el}^0 < V_{el}^{i \neq 0}$  (by definition of the ground state), we obtain a minimum of equation 2.35 when  $|c_0|^2 = 1$  and  $|c_{i \neq 0}|^2 = 0$ , and we also get

$$V_{el}[\psi_e^0] \leq V_{el}[\psi_e^{trial}] \quad \forall \quad \psi_e^{trial} \quad (2.37)$$

This means that whatever wave function we attempt and solve equation 2.25, we will obtain a higher energy than the ground state energy, except if we really have the ground state wave function. This principle is called

the "Variational principle". The criteria (equation 2.37) also holds when the energy is given as a functional of the electron density.

$$V_{el}[\rho^0] \leq V_{el}[\rho^{trial}] \quad \forall \quad \rho^{trial} \quad (2.38)$$

The variational principle gives us a procedure to approximate our ground state wave function or ground state density, by using the calculus of variations<sup>2</sup>. In other words, the approach will be to minimise the energy with respect to the density or the wave function. The wave function that gives the minimum energy will be the ground state wave function, and also be an Eigenfunction of equation 2.15. Note that the trial wave functions will in general not be Eigenfunctions of equation 2.15, but superpositions of the Eigenfunctions, as shown in equation 2.35.

We want to obtain the ground state density and we therefore attempt such a minimisation. However, there are some restrictions on the minimisation. The first constraint, says that it should be possible to obtain the electron density from a wave function as given in equation 2.27. It has been shown that this can be done for all non-negative densities. Secondly we need the total number of electrons to be included and thirdly the density has to arise from an antisymmetric wave function (as electrons are fermions). This can be ensured by having  $\int |\nabla \rho^{\frac{1}{2}}|^2 dr < \infty$  [34]. For the electron density we are left with the problem

$$\begin{aligned} & \text{minimise: } V_{el}[\rho] \\ & \text{with respect to: } \rho \\ & \text{subject to: } \rho(r) > 0 \quad \forall \quad r \\ & \text{subject to: } \int \rho(r) dr = N_e \\ & \text{subject to: } \int |\nabla \rho(r)^{\frac{1}{2}}|^2 dr < \infty \end{aligned} \quad (2.39)$$

The electron density returned from the solution of this minimisation problem gives the ground state electron density, and the energy is the ground state energy. This minimisation draws the outline on how we will obtain both the electron density and the energy.

### 2.3.4 Minimisation of energy

The solution of how to obtain the energy from a given electron density is not trivial. Since we can easily find the energy of noninteracting electrons in a given potential, we choose to use non-interacting electrons as our reference system. We then separate equation 2.32 into a non-interacting part and one correction part as

$$V_{el}[\rho(r)] = T^{NI}[\rho] + V_{ee}^{NI}[\rho] + V_{ext}[\rho] + (T[\rho] - T^{NI}[\rho] + V_{ee}[\rho] - V_{ee}^{NI}[\rho]) \quad (2.40)$$

where the index *NI* stands for "non-interacting". Equation 2.40 is further rewritten to<sup>3</sup>

$$V_{el}[\rho(r)] = T^{NI}[\rho] + V_{ee}^{NI}[\rho] + V_{ext}[\rho] + V_{xc}[\rho] \quad (2.41)$$

where the last term is called the "exchange and correlation" functional. The three first functionals in equation 2.41 are given as

---

<sup>2</sup>Many problems can be formulated as variational problems, e.g. trajectories from classical mechanics can be found by minimising the function  $\int_{t_1}^{t_2} E_{kin}[x(\tau)] - E_{pot}[x(\tau)] d\tau$  with respect to a trajectory  $x(\tau)$ . A good introduction to the calculus of variations can be found in Feynman [37].

<sup>3</sup>By using the famous GEDIOP-method (Gather Everything Difficult into one Problem). Obviously we have here taken two difficult functionals and reduced them into one difficult one, thus reducing the number of complicated things.

$$T^{NI}[\rho^{NI}] = \langle \psi_e^{NI} | -\frac{\hbar^2 \nabla^2}{2m} | \psi_e^{NI} \rangle \quad (2.42)$$

$$V_{ee}^{NI}[\rho^{NI}] = \frac{e^2}{4\pi\epsilon_0} \int \int \frac{\rho(r)^{NI} \rho(r')^{NI}}{|r - r'|} dr dr' \quad (2.43)$$

and

$$V_{ext}[\rho] = -e \int \rho(r) v_{ext}(r) dr \quad (2.44)$$

The last functional  $V_{xc}[\rho]$  will be treated in Chapter 2.3.5. It can be shown that the minimisation (equation 2.39) with respect to  $\rho^{NI}$  gives the same electron density as the minimisation with respect to  $\rho$  [34]. Thus the  $NI$  index in equation 2.43 can be omitted and the true electron density is

$$\rho(r) = \rho(r)^{NI} = \sum_i^{occupied} |\phi_i^{NI}|^2 \quad (2.45)$$

where  $\phi_i^{NI}$  are the non-interacting single electron wave functions if we have found the minima of equation 2.39. This is also why the index  $NI$  was omitted from equation 2.44 in the first place, as this will be the exact interaction energy when the minimisation has been performed.

We can therefore attempt to obtain the non-interaction density in our search for the electron density. This was proposed by Kohn and Sham [38] and the non-interaction density arises from the non-interaction wave functions  $\phi_i^{NI}$  which are called Kohn-Sham orbitals. There exist other approaches to find the ground state density that do not require the construction of orbitals, but orbitals have shown themselves useful in obtaining an accurate kinetic energy and is the method of choice for most practical DFT calculations. The Kohn-Sham orbitals are not actually independent, more pseudo-independent as all the dependent factors are gathered in the exchange-correlation functional  $V_{xc}$ . But as the system of non-interacting electrons gives the same density as the true electron density, we choose to find the density from the non-interacting electrons instead. We solve the non-interaction orbitals by writing equation 2.41 as

$$V_{el}[\rho] = T^{NI}[\rho] + V_{eff}[\rho] \quad (2.46)$$

We have now split our energy into one kinetic energy part  $T^{NI}$  and one pseudo-potential energy part  $V_{eff}$  (pseudo-potential because it also contains a part of the kinetic energy  $T - T^{NI}$ ). We choose further to consider the pseudo-potential as a real potential, which enables us to construct a Hamilton operator as

$$\hat{H}_e = \hat{T}^{NI} + \hat{V}_{eff} \quad (2.47)$$

The potential operator  $\hat{V}_{eff}(r)$  is not trivial to construct, but will be

$$\hat{V}_{eff} = -\frac{1}{e} \frac{\delta V_{eff}[\rho]}{\delta \rho} = -\int \frac{e\rho(r')}{4\pi\epsilon_0|r - r'|} dr' + v_{ext}(r) - \frac{1}{e} \frac{\delta V_{xc}[\rho]}{\delta \rho} \quad (2.48)$$

The electron density is then obtained by solving the Schrödinger equation (equation 2.15) for each electron independently in the potential  $V_{eff}$ . Thus we solve

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + \hat{V}_{eff}\right] \phi_i(r) = \epsilon_i \phi_i(r) \quad (2.49)$$

to get the single electron orbitals  $\phi_i$ . These equations 2.49 are called the Kohn-Sham equations. Note that  $V_{eff}$  is dependent on the orbitals obtained (through the electron density), and thus we need to iterate until

we get a self-consistent solution.

The iteration procedure normally goes as:

1. Guess  $\rho$
2. Calculate  $\hat{V}_{eff}$  from equation 2.48
3. Solve equations 2.49 to obtain  $\phi_i$ 's
4. Calculate a new  $\rho$  from equation 2.45
5. Repeat step 2-5 until self-consistency

How to construct the  $\hat{V}_{eff}$  is treated in Chapter 2.3.5 and how to solve the Kohn-Sham equations is treated in Chapter 2.3.10.

### 2.3.5 Exchange and correlation

The challenge in density functional theory is the estimation of the exchange-correlation functional  $V_{xc}$ . There are numerous approaches to this, and this report will only describe the functional employed in this project, namely the B3LYP functional.

The B3LYP functional was developed from the hybrid orbital type of Becke[39], using the Local Density Approximation (LDA) for exchange, mixed with exact exchange and the Becke88[40] functional for exchange correction, and the VWN functional from Vosko et al.[41] as a LDA approximation of correlation together with the LYP functional from Lee et al.[42] for correlation correction. This combination was first made by Stevens et al.[43]. Some different implementations of the VWN functional have been used, and for a review of the B3LYP functional the article by Hertwig and Koch[44] is recommended<sup>4</sup>.

The B3LYP functional is constructed as

$$V_{xc}^{B3LYP}[\rho] = (1 - a_0)E_x^{LSDA} + a_0E_x^{exact} + a_xE_x^{B88} + a_cE_c^{LYP} + (1 - a_c)E_c^{VWN} \quad (2.50)$$

where  $a_0$ ,  $a_x$  and  $a_c$  are coefficients fitted to a number of experimental data and found to be 0.2, 0.72 and 0.81 respectively. The different contributing terms will be described below.

### 2.3.6 Exact exchange

The exact exchange is given as a function of the Kohn-Sham orbitals, by introducing an antisymmetrized wave function as a Slater determinant.

$$\psi_{Slater} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(r_1) & \phi_2(r_1) & \cdots & \phi_N(r_1) \\ \phi_1(r_2) & \phi_2(r_2) & \cdots & \phi_N(r_2) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_1(r_N) & \phi_2(r_N) & \cdots & \phi_N(r_N) \end{vmatrix} \quad (2.51)$$

This would be the exact antisymmetrized wave function for non-interacting electrons. We are using non-interacting electrons as our reference system, thus this will give the exact exchange. Then by solving the Schrödinger equation (equation 2.15) we get the energy including  $E_x^{exact}$ , and the additional term compared to solving equation 2.49 is the exchange term, which is given as

<sup>4</sup>The history of B3LYP is a bit confusing, and a discussion on this can be found on the Computational Chemistry List. <http://www.ccl.net/chemistry/resources/messages/2002/05/22.008-dir/index.html>

$$E_x^{exact} = \frac{1}{2} \sum_{jk} \int \int \frac{e^2 \phi_j^*(r) \phi_k^*(r') \phi_j(r') \phi_k(r)}{4\pi\epsilon_0 |r - r'|} \quad (2.52)$$

It is tempting to separate the exchange and correlation in two separate parts such that

$$E_{xc} = E_x + E_c \quad (2.53)$$

since we can calculate the exchange energy  $E_x$  exactly. However this has not been successful since they are closely connected and it is easier to estimate both effects at the same time. A good argument for keeping a part of the exact exchange energy is given by Becke[45].

### 2.3.7 Local Density Approximation

For a homogenous electron gas it has been found that the energy density of exchange is[34]

$$\epsilon_x = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{\frac{1}{3}} \rho^{\frac{1}{3}} \quad (2.54)$$

in atomic units. By assuming that this equation holds for inhomogenous electron densities, we obtain the exchange energy as

$$E_x^{LSDA} = \int \rho(r) \epsilon_x dr \quad (2.55)$$

This approximation assumes that the electron density is locally homogenous, thus the approximation is denoted as "Local Spin Density Approximation" or "LSDA" for the spin case and "Local Density Approximation" or "LDA" for the non-spin case.

Similarly, for the correlation energy we can calculate numerically the correlation energy density for different densities of homogenous electron gas. A fit to these data has been found and is given as[41]

$$\begin{aligned} \epsilon_c^{VWN} = & A \left( \ln \frac{x^2}{X} + \frac{2b}{Q} \tan^{-1} \frac{Q}{2x+b} \right. \\ & \left. - \frac{bx_0}{X(x_0)} \left[ \ln \frac{(x-x_0)^2}{X(x)} + \frac{2(b+2x_0)}{Q} \tan^{-1} \frac{Q}{2x+b} \right] \right) \end{aligned} \quad (2.56)$$

in atomic units, where  $X(x) = x^2 + bx + c$ ,  $x = r_s^{\frac{1}{2}} Q = (4c - b^2)^{\frac{1}{2}}$ , and  $x_0 = -0.40926$ ,  $b = 13.0720$  and  $c = 42.7198$  are some parameters and  $r_s$  is the radius of a sphere that contains one electron, thus  $\frac{1}{\rho} = \frac{4\pi r_s^3}{3}$ . Similarly to equation 2.55, the total functional is

$$E_c^{VWN} = \int \rho(r) \epsilon_c^{VWN} dr \quad (2.57)$$

This functional for correlation was proposed by Vosko, Wilk and Nusair[41] (denoted expression V in the original article) and also contain an interpolation for spin-systems. This interpolation formula can be found in the original article. An alternative interpolation, based on a Random Phase Approximation (RPA)[46] has also been denoted the VWN-functional, and arises from expression III in the article by Vosko et al[41]. This is the functional implemented in the B3LYP version of the quantum chemistry programs Gaussian[47] and Jaguar[48].

### 2.3.8 Becke's exchange correction

The LSDA-approximation to the exchange energy undershoots the proper exchange energy for molecules by approximately 10%[40]. A typical way of making the functional less local is to include some effect of the gradient of the density, a method called Generalized Gradient Approximation (GGA). Such a correction term to the exchange energy was therefore proposed by Becke[40] as

$$E_x^{B88} = E_x^{LSDA} - \beta \int \frac{\rho^{\frac{4}{3}}}{2^{\frac{1}{3}} 1 + 6\beta x \sinh^{-1} x} x^2 d^3r \quad (2.58)$$

in atomic units, where  $x = \frac{2^{1/3}|\nabla\rho|}{\rho^{\frac{4}{3}}}$  and  $\beta$  is a parameter fitted to be  $\beta = 0.0042$  a.u.

### 2.3.9 LYP correlation correction

Also the LDA for the correlation is not sufficiently accurate, and we choose to include some effect of the gradient here also. A GGA formula containing both the gradient and the second derivative of the density was given by Lee, Young and Parr[42] by using a correlation function of the Colle-Salvetti form[49], and is given as

$$E_c^{LYP} = -a \int \frac{1}{1 + d\rho^{-\frac{1}{3}}} (\rho + b\rho^{-\frac{2}{3}} [C_F \rho^{\frac{5}{3}} - 2t_w + (\frac{1}{9}t_w + \frac{1}{18}\nabla^2\rho)] e^{-c\rho^{-\frac{1}{3}}}) dr \quad (2.59)$$

in atomic units, where

$$t_w = \frac{1}{8} \frac{|\nabla\rho|^2}{\rho} - \frac{1}{8} \nabla^2\rho \quad (2.60)$$

and  $C_F = \frac{3}{10}(3\pi^2)^{\frac{2}{3}}$ , and  $a = 0.049$ ,  $b = 0.132$ ,  $c = 0.2533$  and  $d = 0.349$  are fitted parameters.

We can now construct the B3LYP functional and perform the minimisation as described in Chapter 2.3.4

### 2.3.10 Solving the Kohn-Sham equations

The Kohn-Sham equations (equations 2.49) have to be solved during the iteration procedure as described in Chapter 2.3.4 and we need to construct the orbitals  $\phi_i$ . These orbitals are in general not given as analytical functions, but in the same manner as equation 2.34, we can write the orbitals as a linear combination of linearly independent wave functions  $f_i(r)$

$$\phi_i(r) = \sum_{i=1}^{\infty} c_i f_i(r) \quad (2.61)$$

as long as the  $f_i(r)$ 's form a complete set, which will require an infinite sum. Optimising an infinite number of  $c_i$ 's due to the minimisation problem (equation 2.39), will take an infinitely long time, so we will take a limited set of basis functions  $f_i(r)$  in our expansion, thus introducing an approximation. These sets of functions will be described in Chapter 2.4.

We combine equation 2.49 and 2.61 and get

$$[-\frac{\hbar^2\nabla^2}{2m} + \hat{V}_{eff}]\phi_i(r) = \hat{F}\phi_i(r) = \hat{F}\sum_j c_j f_j = \epsilon_i \sum_j c_j f_j \quad (2.62)$$

where  $\hat{F} = -\frac{\hbar^2\nabla^2}{2m} + \hat{V}_{eff}$ . We want to find the coefficients  $c_j$  such that this equation is correct, and this we attempt in an iterative manner. We take an initial set of  $c_j$ 's, calculate the left side of equation 2.62

and estimate the new  $c_j$ 's as the result on the right side of equation 2.62. After some steps, the equality in equation 2.62 is fulfilled and we have the correct  $c_j$ 's. However, it is difficult to separate the  $c_j$ 's from equation 2.62 and we do a trick of multiplying with  $\phi_i(r)$  and integrate over all space obtaining

$$\langle \phi_i(r) | \hat{F} | \phi_i(r) \rangle = \langle \sum_j c_j f_j | \hat{F} | \sum_j c_j f_j \rangle = \langle \sum_j c_j f_j | \epsilon_i \sum_j c_j f_j \rangle \quad (2.63)$$

This equation we recast into

$$c^T F c = \epsilon_i c^T S c \quad (2.64)$$

where  $F_{ij} = \langle f_i | \hat{F} | f_j \rangle$ ,  $S_{ij} = \langle f_i | f_j \rangle$  and  $c = [c_1 \dots c_N]^T$ . This will be equivalent to

$$F c = \epsilon_i S c \quad (2.65)$$

If we orthonormalise the  $f_i$ 's, the  $S$  matrix will be the identity matrix, and we will obtain an equation of the type

$$F' c' = \epsilon_i c' \quad (2.66)$$

The  $F'$  matrix is calculated from the  $\langle f_i | \hat{F} | f_j \rangle$  integrals by using the density of the first guess. A new set of coefficients  $c_j$  can then be found as the Eigenvectors of  $F'$  from equation 2.66. This is then repeated by calculating a new density, obtaining new coefficients and so on, as described in Chapter 2.3.4. All the 1-electron orbitals are defined by equation 2.49, and we therefore search for the  $N_e$  Eigenfunctions  $\phi_i$  with the  $N_e$  lowest Eigenvalues.

The integrals  $\langle f_i | \hat{F} | f_j \rangle$  can normally not be solved analytically and have to be performed numerically, thus we need to specify a grid. The integration will only be exact for an infinitely dense grid, but a normal approximation is to use 1000-10000 grid points for each atom[32].

## 2.4 Basis sets

As explained in Chapter 2.3.10, we need to form a set of linearly independent wave functions to use as a basis for the wave function. To attempt to minimise the number of basis functions, we try to form functions similar to the analytical solutions to one-electron systems. Such a procedure has been performed in many ways, e.g. using Slater Type Orbitals (STO's)[50]. A more common practice is to use Gaussian Type Orbitals (GTO's), which are on the form

$$f(x, y, z) = N x^{l_x} y^{l_y} z^{l_z} e^{-ar^2} \quad (2.67)$$

where the  $l_i$ 's are integers. To approximate functions of the STO-type (which have the shape of non-interacting 1 electron orbitals), we do a linear combination of a certain number of GTO's, often called a "contraction". A typical set of such STO-like orbitals are the STO-nG basis sets[51], where a STO has been approximated by n GTO's. Another common approach is to separate valence orbitals and core orbitals, as the valence orbitals describe most of the chemical properties of a molecule. There are many such "split-valence" orbitals, and the one applied in this project is described below.

Another factor to consider is the placement of the orbitals. In principle, we could center the orbitals wherever we want, but almost all implementations center the orbitals on the nuclei, as this is the case for 1-electron systems without external fields.

### 2.4.1 6-311++G(d,p) basis set

The 6-311G basis set[52, 53] describe the core orbitals as one contraction of 6 GTO's, and the valence orbitals as one contraction of 3 GTO's and 2 uncontracted GTO's, thus the name 6-311. More specifically for 1<sup>st</sup> row atoms, the core 1s-orbital is given as one contraction of 6 GTO's, the valence 2s-orbital and the 3 2p-orbitals are given as linear combinations of 3,1 and 1 contracted GTO's. The 1s orbital of hydrogen is also described as a linear combination of 3,1 and 1 contracted GTO's. This is described in detail in the original paper[52].

This set of contracted orbitals is optimised to describe the proper orbitals. However some effects are not described properly. The contracted set is not flexible enough to polarize properly, and we therefore often add such flexibility by adding another GTO for each atom. The "(d,p)" in 6-311++G(d,p) means that 5 d-type GTO's (1 per d-orbital direction) have been added to the heavy atoms, while 3 p-type GTO's (1 per p-orbital direction) have been added to hydrogen. Also f-type orbitals can be added as polarization orbitals.

Another problem with the contracted basis sets is that the constructed wave function approaches zero earlier than the true wave function for long distances from the atomic centre. In particular anions[54, 55], lone-pair electrons and weakly bounded systems such as hydrogen-bond systems[56] suffer from this poor description. A correction for describing such systems properly is then included by adding a flat and broad s-type GTO, also called a "diffuse" function. Such a function can be added only to the heavy atoms (signified by one +, e.g. 6-311+G(d,p)) or also to hydrogens (signified by two +'s, e.g. 6-311++G(d,p)).

### 2.4.2 Basis Set Superposition Error

Since we are operating with a limited set of wave functions, we will obtain approximate results. To attempt to zero out the error, it is important to have the same amount of basis functions available for systems that we want to compare. The normal approach is to center the basis functions on the nuclei, but this gives a problem when calculating bond energies. The bond energy for  $A - B$  will be the energy difference between

1.  $A$  and  $B$  bonded
2.  $A$  and  $B$  separated by a large distance

If we look at situation 1 the electrons on atom  $A$  have effectively more available basis functions, since it can also use basis functions from the nearby atom  $B$ . In situation 2 atom  $B$  is too far away to share its basis functions with atom  $A$ , thus atom  $A$  experiences less available basis functions. By the variational principle (Chapter 2.3.3) this will lead to an unphysical stabilisation of the bond, an effect called the "Basis Set Superposition Error" BSSE. In many cases this error is large, and we need to correct it somehow. A common approach is the Counterpoise (CP) method[57]. This correction is employed by creating orbitals also in the unbonded state. E.g. to get the corrected bond energy, we have to compute the energy of the  $A - B$  complex, and subtracting the energy of  $A - X_B$  and  $X_A - B$ , where  $X_A$  and  $X_B$  is symbolising the orbitals of  $A$  and  $B$  without any nuclei or electrons, often called "ghost" orbitals. Then the electrons of the unbonded  $A$  and  $B$  will have the same number of basis functions as in the bounded state, and the energies can be compared.

The CP-method works well when we compare atoms at some fixed distance. However the energy correction will also affect the geometry of the system. The normal CP-correction is performed only at pre-optimised structures optimised without any BSSE correction, but also methods that optimise the structure and simultaneously correct for BSSE are available[58].

### 2.4.3 Choice of method and basis set

The choice of method in this project has been a DFT method (Chapter 2.3.1), with the functional B3LYP (Chapter 2.50). DFT-methods give a high accuracy and a low cost compared to wave function methods. The B3LYP functional is one of the most versatile and accurate of the DFT functionals, and has been applied to many problems. It is also implemented in most available programs for quantum chemistry calculations. The accuracy of this functional has been reported to give an average deviation in the energy prediction of 3.11 kcal/mol on the 148-molecules test set G2-2[59] by comparing to experimental energies. This deviation is much higher than the deviation obtained by using G2-theory[60] which gives an average energy deviation of 1.58 kcal/mol. Parts of this set have been used for parametrisation of the B3LYP functional[39], thus a comparison using bigger test sets will give a better answer to the accuracy. Such a comparison has been performed by Boese and Handy[61] reporting an average deviation of 6.3 kcal/mol of B3LYP predicted energies to experiment, and by Curtiss et al. [62] who report an average energy deviation of 4.14 kcal/mol. Curtiss et al. also report some problems with the B3LYP functional for larger molecules e.g. large  $sp^3$  hydrocarbons and especially large cyclic  $sp^3$  hydrocarbons, and this is assumed to be a result of the ability of B3LYP to describe C-C bonds[62]. The B3LYP functional also seems to be able to describe radical reactions as Bernardi et al.[63] report that the unrestricted B3LYP functional gives errors between 9 and 18 % on reaction barriers for open-shell systems consisting of small hydrocarbon radicals and is superior to unrestricted HF and MP2 methods.

Other DFT-functionals exist of similar accuracy as the B3LYP, but B3LYP has been chosen as it is commonly used for calculations, including calculations on carbohydrates[64, 65, 66, 67] and force field development[68] and has been used for the previous ReaxFF parametrisations[6].

The accuracy of our results is dependent on the basis set available. A limited number of basis functions will in most cases give an approximation to the real wave function, thus leading to errors. However, an increased number of basis functions will lead to increased computational time, and we have to settle for a trade-off between computational time and accuracy.

The basis set chosen for the calculations is the 6-311++G(d,p)[52, 53], described in Chapter 2.4.1. This is similar to what has been used for other force field developments[68], and also what has been used for some recent ReaxFF parametrisations[69, 70]. A basis set for carbohydrates should include diffuse functions on heavy atoms[71, 65, 72] and Csonka[71] suggests the use of 6-31+G(d,p) or 6-311+G(d,p) in combination with B3LYP, and at the same time claims that diffuse functions on hydrogen are not necessary. Lii et al.[65] show that the BSSE is much lower when diffuse functions are introduced, e.g. for a water dimer the BSSE is approximately 7% of the total bond energy for the combination B3LYP/6-311++G(2d,2p), while for the combination B3LYP/6-31G(d,p) is reported to be about 32 %. Van der Vaart and Merz[73] show that the BSSE for hydrogen bonded clusters is of the order 5% with the combination B3LYP/6-311++G(d,p), much lower than the BSSE calculated for smaller basis sets and for MP2[74] methods.

Given the low BSSE given by the basis set, that the accuracy of B3LYP already is not perfect, and that the data will be used for force field parametrisation (which will probably be even less accurate), we also choose to neglect the BSSE and proceed with uncorrected results.

## 3 Force Fields

When we employed the Born-Oppenheimer approximation, we had to solve equation 2.23 to obtain the total energy of our system as a function of the nuclear coordinates. As quantum chemistry calculations are only computationally feasible for small systems, we would like to approximate the electronic energy from

this equation by simpler functions that can be evaluated easier, such that we can do calculations on larger systems and for more steps.

### 3.1 Common force field terms

For conventional non-reactive force fields (like GROMOS [2], AMBER [3], GROMACS [4] and NAMD [5]) we normally approximate the system energy by considering two main contributions, one for energy involved with bonded atoms and one for non-bonded interactions.

$$E_{system} = E_{nonbond} + E_{bond} \quad (3.1)$$

#### 3.1.1 Non-bonded interactions

The non-bonding energy can be split into 2 parts

1. Classical coulombic energy  $E_c$
2. Van der Waals energy  $E_{vdw}$

The classical coulombic energy describes the interaction between charges. We want to describe the interactions between charge clouds, and this we can do by approximating the charge distributions by point charges, point dipoles and higher point multipoles. The interaction energy will then be the interactions between all multipoles on atom A and all multipoles on atom B, thus leading to a high number of terms. Normally we only keep the point charges, and in some cases also the point dipoles for some polarizable force fields[75]. Here we only consider interactions between point charges, which are given by Coulomb's formula

$$E_c = \frac{q_i q_j}{4\pi\epsilon_0 |r_i - r_j|} \quad (3.2)$$

where  $q_i$  denotes the charge of each particle. The charge of each atom is normally employed as an effective charge, thus not being the total charge of an atom or the nucleic charge, but rather parametrised after some data.

The van der Waals energy contains a repulsive term at short distances. Because we have approximated our charge distribution by a point charge, we get the wrong interaction at short distances as the real charge distributions start to overlap, which creates a repulsive force due to the coulombic energy (equation 3.2). In addition, the electrons are more repulsive than the classical Coulomb energy (equation 3.2) can describe. This is because electrons are fermions and two electrons with equal spin cannot exist in the same energy state, thus increasing their repulsivity. This is the quantum effect which is called "exchange". The repulsive term of the van der Waals forces corrects for the additional Coulomb and exchange energy that an overlap would have given. At longer distances, the forces become attractive due to temporary dipoles induced in the two interacting electron clouds<sup>5</sup>. These attractive forces are often referred to as London dispersion forces.

The energy term arising from van der Waals forces can be approximated by a Morse potential, thus

$$E_{vdw} = D[1 - e^{\alpha(r - r_{vdw}^0)}]^2 \quad (3.3)$$

where  $D > 0$ ,  $\alpha < 0$  and  $r_{vdw}^0 > 0$  are fitted parameters.

---

<sup>5</sup>This effect is employed by geckos when they climb on your wall. That's why they can climb almost all surfaces, wet or dry, even in vacuum, but not on your teflon frying pan, which gives too low van der Waals forces (K. Autumn, American Scientist 2006, vol 94, number 2)

### 3.1.2 Bonded interactions

Non-reactive force fields normally deal with another set of equations for atoms that are bonded to each other. We separate the interactions in different terms, a covalent bond interaction energy, a valence angle energy and a torsion angle energy as

$$E_{bond} = E_{cov} + E_{val} + E_{tors} \quad (3.4)$$

The covalent bond is not suppose to break, and can then be restrained by a simple harmonic potential as

$$E_{cov} = \frac{k}{2}(r - r_0)^2 \quad (3.5)$$

where  $k > 0$  is the force constant,  $r_0 > 0$  a reference bond length, and  $r$  the actual bond length. We see that the formulation of the bond energy in equation 3.5 does not allow bond breaking, as the bond energy would go to infinity for long bond lengths. The valence angle is the angle  $A - B - C$  and is for a non-reactive force field well defined by the number of lone pairs of the central atom and its bonding type (double bond, triple bond). This angle is also simply restrained by a harmonic potential as

$$E_{val} = \frac{k}{2}(\theta - \theta_0)^2 \quad (3.6)$$

where  $k > 0$  is the force constant, and  $\theta_0$  a reference angle.

The torsion angle represents the dihedral angle  $A - B - C - D$ , which is the angle between the two planes intersected by atoms  $A - B - C$  and atoms  $B - C - D$  respectively. The torsion angle energy arises from the restrictions on rotation around bonds. This could be because of double bond character or steric hindrance of the subgroups, favouring "anti" configuration. The torsion angle however is allowed to vary more freely than the valence angle, and can obtain all values between 0 and  $2\pi$ . There are multiple minimas and maximas for these angles, and we can approximate the energy with trigonometric functions (taking some terms from a fourier series) and get

$$E_{tor} = \sum_i k_i(1 - i \cos \phi) \quad (3.7)$$

where the  $k_i$ 's are constants.

## 3.2 Reactive Force Fields

A non-reactive force field has the advantage that it knows the topology of the molecules (which atoms that are bonded to each other) and also the properties of each bond (if it is a double bond, has a lone pair, etc.). For a reactive force field this changes over time, and the force field somehow has to know which and what kind of bonds are present in order to get the energy of the system. A method to deal with this issue has been developed by the introduction of a bond order function. Bond order is a concept mostly used for describing the number of bonds between a pair of atoms, e.g. single, double and triple bonds. The bond order function calculates the bond order of a bond to a value between 0 and 3, and simultaneously scales the bond interactions to range between no bond and a triple bond. This enables us to treat all interactions between atoms with the same set of functions, and also in principle allows all atoms to be non-bonded, single bonded, double bonded and triple bonded. The bond order function is also a continous function, thus we will have a continous transition from a single bonded structure to a double bonded one. The introduction of the bond order function was first done by Abell[19] who showed that the bond energy of a system of atoms could be approximated as

$$E_{bonds} = \sum_i \sum_j^{nb(i)} [V^R(r_{ij}) - b_{ij}V^A(r_{ij})] \quad (3.8)$$

where  $nb(i)$  denotes the neighbours of atom  $i$ ,  $V^R(r_{ij})$  is the repulsion energy,  $V^A(r_{ij})$  is the attraction energy and  $b_{ij}$  a function similar to the bond order. The function  $b_{ij}$  scales the energy, and was by Abell found to depend on the number of atoms in the first shell for metals, thus dependent on the coordination of the atom. This formulation of the energy enabled the description of bond energy for different packing in metal structures. To relate  $b_{ij}$  to the bond order, we define  $BO_{ij}$  as a scaled version of  $b_{ij}$  obtaining the values between 0 and 3.

Different formulations of the bond order have been described by Tersoff[20] and Brenner[21, 23]. Tersoff defined the bond order as

$$BO_{ij}^{Tersoff} = f(r_{ik}, r_{jm}, \theta_{ijk}, \theta_{jim}) \quad (3.9)$$

thus a function of the bond distance  $r$  and the bond angle  $\theta$  to the neighbours of atom  $i$  and  $j$ . Brenner defines the bond order similarly, but includes the number of neighbouring atoms as

$$BO_{ij}^{Brenner} = f(r_{ik}, r_{jm}, \theta_{ijk}, \theta_{jim}, N^a, N^b, \dots) \quad (3.10)$$

where  $N^a$  is the number of neighbouring atoms of type  $a$ . The Brenner potential was developed for hydrocarbons, thus the bond order only includes the number of hydrogen and carbon atoms. ReaxFF on the other hand defines the bond order as

$$BO_{ij}^{ReaxFF} = f(r_{ij}, r_{ik}, r_{jm}) \quad (3.11)$$

The non-reactive force fields contain a number of energy terms for bonded interactions (Chapter 3.1.2) and all these have to be scaled as a function of the bond order to handle bond breaking.

### 3.3 ReaxFF Force Field Terms

ReaxFF allows bonds to break and form, and therefore has to modify the energy terms from Chapter 3.1. ReaxFF deals with this bond breaking by using the bond order  $BO$  as described in Chapter 3.2. The main contributions to the force field are energy terms of the type included in non-reactive force fields, that are scaled as a function of the bond order, but some correction terms for special cases (e.g. conjugated systems) are also included. The number of energy terms included in ReaxFF is quite substantial, and only the  $BO$  function (Chapter 3.3.1) and the bond energy where the  $BO$  function is employed (Chapter 3.3.2) will be described in this report. This demonstrates the use of the  $BO$  function and similar procedures are applied to all bonded interactions. A full description of the energy terms ReaxFF applies can be found as supporting information to the article by Nielson et al.[25], and also in appendix A.

#### 3.3.1 Bond Order

The bond order function is a function which tells us what kind of bonding we have. It is a continuous variable between 0 and 3, where  $BO = 1$ ,  $BO = 2$  and  $BO = 3$  corresponds to a  $\sigma$ -bond, a double bond, and a triple bond. When  $BO = 0$  there is no bond at all. The bond order is given by ReaxFF as

$$BO_{ij} = BO_{ij}^{\sigma} + BO_{ij}^{\pi} + BO_{ij}^{\pi\pi} = e^{p_{bo,1}(\frac{r_{ij}}{r_{\sigma}})^{p_{bo,2}}} + e^{p_{bo,3}(\frac{r_{ij}}{r_{\pi}})^{p_{bo,4}}} + e^{p_{bo,5}(\frac{r_{ij}}{r_{\pi\pi}})^{p_{bo,6}}} \quad (3.12)$$

where  $p_{bo,1}, p_{bo,3}, p_{bo,5} < 0$  and  $p_{bo,2}, p_{bo,4}, p_{bo,6} > 0$  are parameters, and  $r_{\sigma}, r_{\pi}$  and  $r_{\pi\pi}$  are reference bond lengths. By this definition, the bond order is only a function of the bond distance. To demonstrate how the

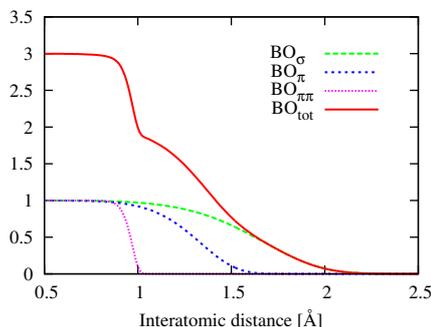


Figure 1: Bond order for the O-O bond as a function of O-O distance

bond order function changes as a function of the bond distance, the bond order for a O-O bond is plotted in Figure 1. We see from the figure that by using equation 3.12 for the bond order, we get a triple bond to be present for O-O distances less than 1 Å, a double bond for a bond distance of 1 Å, a single bond for a bond distance of 1.5 Å, and no bond for bond distances greater than 2 Å. As explained in Chapter 3.2, all atoms are in principle allowed to participate in all types of bonds, so the bond order itself does not tell us which type of bond that is the energetically most stable.

The total bond order for one atom will be the sum of the bond orders from the bonds which it participates in. This sum should be equal to the valence of the atom, in other words, a hydrogen should only be bonded to 1 atom, while carbon can be bonded to 4. The bond order definition of equation 3.12 typically gives a too high total bond order for an atom, and ReaxFF therefore introduces a correction procedure. We first define

$$\Delta_i = \sum_{j=1}^{nb(i)} BO_{ij} - Val_i \quad (3.13)$$

to be the overcoordination of each atom, where  $Val_i$  is the valence number (number of bonds it should form) of the atom in question. We want to adjust this overcoordination close to 0 to make the sum of bond orders equal to the valence number. Some atoms can form more bonds, depending on the molecule in question, and also lone pairs can give some bonding interactions with other atoms. ReaxFF therefore introduces a second overcoordination parameter

$$\Delta_i^{ov} = \sum_{j=1}^{nb(i)} BO_{ij} - Val_i^{ov} \quad (3.14)$$

where  $Val_i^{ov}$  is the sum of the valence and the number of lone pairs for each atom, thus giving 4 for Nitrogen and 4 for Oxygen (and also the value 4 for all atoms that seeks to fill their outer shell with 8 electrons). ReaxFF employs these overcoordination parameters to correct the bond order by the following set of equations

$$\begin{aligned} BO_{ij}^{\sigma,corr} &= f_1(\Delta_i, \Delta_j) f_4(\Delta_i^{ov}, BO_{ij}) f_5(\Delta_j^{ov}, BO_{ij}) BO_{ij}^{\sigma} \\ BO_{ij}^{\pi,corr} &= [f_1(\Delta_i, \Delta_j)]^2 f_4(\Delta_i^{ov}, BO_{ij}) f_5(\Delta_j^{ov}, BO_{ij}) BO_{ij}^{\pi} \\ BO_{ij}^{\pi\pi,corr} &= [f_1(\Delta_i, \Delta_j)]^2 f_4(\Delta_i^{ov}, BO_{ij}) f_5(\Delta_j^{ov}, BO_{ij}) BO_{ij}^{\pi\pi} \end{aligned} \quad (3.15)$$

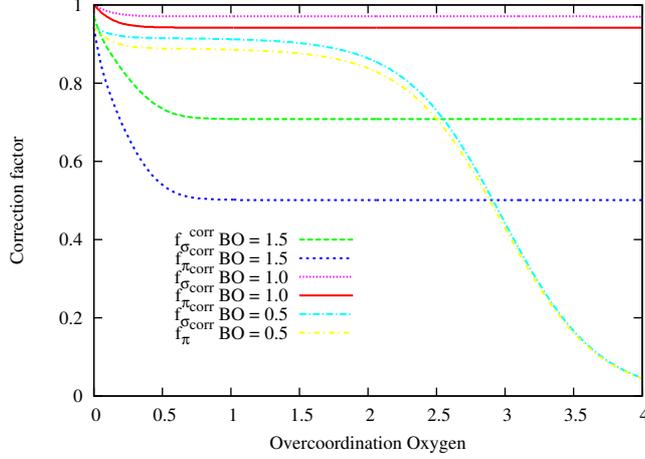


Figure 2: The correction factors for the bond order of a O-H bond as a function of the overcoordination  $\Delta_O$ .  $\Delta_H = 0$  for  $BO = 0.5$  and  $1$  and  $\Delta_H = 0.5$  when  $BO = 1.5$ .

$$f_1(\Delta_i, \Delta_j) = \frac{1}{2} \left[ \frac{Val_i + f_2(\Delta_i, \Delta_j)}{Val_i + f_2(\Delta_i, \Delta_j) + f_3(\Delta_i, \Delta_j)} + \frac{Val_j + f_2(\Delta_i, \Delta_j)}{Val_j + f_2(\Delta_i, \Delta_j) + f_3(\Delta_i, \Delta_j)} \right] \quad (3.16)$$

$$f_2(\Delta_i, \Delta_j) = e^{-p_{boc,1}\Delta_i} + e^{-p_{boc,1}\Delta_j} \quad (3.17)$$

$$f_3(\Delta_i, \Delta_j) = -\frac{1}{p_{boc,2}} \ln \left[ \frac{e^{-p_{boc,2}\Delta_i} + e^{-p_{boc,2}\Delta_j}}{2} \right] \quad (3.18)$$

$$f_4(\Delta_i^{ov}, BO_{ji}) = \frac{1}{1 + e^{p_{boc,5} - p_{boc,3}[p_{boc,4}BO_{ij}^2 - \Delta_i^{ov}]} \quad (3.19)$$

$$f_5(\Delta_j^{ov}, BO_{ji}) = \frac{1}{1 + e^{p_{boc,5} - p_{boc,3}[p_{boc,4}BO_{ij}^2 - \Delta_j^{ov}]} \quad (3.20)$$

where  $p_{boc,i} > 0$  are parameters.

These equations are perhaps a bit difficult to interpret, and to demonstrate how they work, the bond order correction factor for a O-H bond has been plotted for different total bond orders of the oxygen atom as a function of the overcoordination in Figure 2. We see that if there is no overcoordination on the oxygen, the correction factors are 1. If the hydrogen is loosely bonded, thus giving a bond order of 0.5, its bond order is quenched if the overcoordination of oxygen bypasses 2. This is a result of the lone pair on oxygen which allows this overcoordination. If the hydrogen is tightly bound (with a bond order of 1.5), the bond correction is dominated by the overcoordination on hydrogen, thus scaling the bond order with a factor 0.5 for the  $\pi$ -bond contributions and 0.7 for  $\sigma$ -bond contributions. However, this effect is reduced if the overcoordination of oxygen is small, since then the oxygen doesn't want to reduce its coordination as much. The initial step in the calculation of bond order (equation 3.12) only uses the bond distance, but the correction procedure includes the effect from bonds to neighbouring atoms, thus giving in total the bond order as a function of the type in equation 3.11.

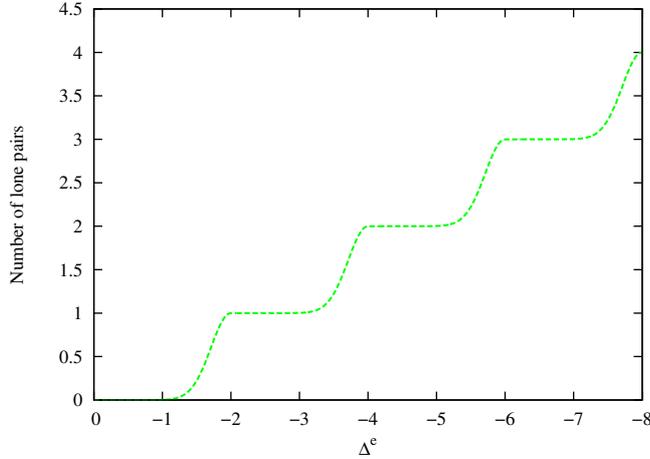


Figure 3: Continuous definition of lone pairs

### 3.3.2 Bond Energy

When the bond order has been established, we can calculate the different energy terms. The bond energy is given by

$$E_{bond} = -D_e^\sigma BO_{ij}^\sigma e^{p_{be,1}(1-(BO_{ij}^\sigma)^{p_{be,2}})} - D_e^\pi BO_{ij}^\pi - D_e^{\pi\pi} BO_{ij}^{\pi\pi} \quad (3.21)$$

where the  $D_e^i$ 's and  $p_{be,i}$ 's are parameters. For systems where we still have overcoordination after the bond order correction (e.g. if carbon really has made 5 bonds), we introduce some extra energy terms. This is to push the system away from unrealistic overcoordinated systems. The overcoordination can arise from the presence of lone pairs and to account for lone pairs, we introduce an energy term that is dependent on the number of lone pairs. The number of lone pairs can be found by the equation

$$n_{lp,i} = -\text{int}\left(\frac{\Delta_i^e}{2}\right) + e^{-p_{lp,1}[2+\Delta_i^e-2\text{int}(\frac{\Delta_i^e}{2})]^2} \quad (3.22)$$

where the  $\text{int}()$  function removes any decimals, (rounding to nearest integer in the direction of 0),  $p_{lp} > 0$  is a parameter and  $\Delta_e$  is defined as

$$\Delta_i^e = \sum_{j=1}^{nb(i)} BO_{ij} - N_i^{eval} \quad (3.23)$$

where  $N_i^{eval}$  is the number of valence electrons for the atom. To show the continuous definition of lone pairs, the number of lone pairs as a function of  $\Delta_i^e$  is plotted in Figure 3. We want the number of lone pairs to be equal to the correct number of lone pairs for each atom (e.g. 2 for oxygen, 1 for nitrogen), and we assign a penalty energy if  $n_{lp,i}$  from equation 3.22 deviates from this. This penalty energy is given as

$$E_{lp} = \frac{p_{lp,2}\Delta_i^{lp}}{1 + e^{-75\Delta_i^{lp}}} \quad (3.24)$$

where  $p_{lp,2} > 0$  is a parameter and

$$\Delta_i^{lp} = n_{lp,i}^0 - n_{lp,i} \quad (3.25)$$

where  $n_{lp,i}^0$  is the correct number of lone pairs of the atom. This energy penalty dominates when  $\Delta_i^{lp} > 0$ , thus when the number of lone pairs is too small. This energy can be considered as an energy of keeping the lone pair together and prevent the electrons to participate in bonds, and it will also to some extent prevent overcoordination (as this will lead to a reduction of lone pairs). However, by looking back at Figure 3 we see that this energy-term also will go approximately stepwise, and we therefore introduce another energy term which is a function of the overcoordination.

We allow the atom to be overcoordinated if the lone pairs have been split up, thus reducing the overcoordination penalty when the lone pairs have been split. We define another overcoordination as

$$\Delta_i^{lpc} = \Delta_i - \frac{\Delta_i^{lp}}{1 + p_{ov,3} e^{p_{ov,4} [\sum_j^{nb(i)} (\Delta_j - \Delta_j^{lp}) (BO_{ij}^\pi + BO_{ij}^{\pi\pi})]}} \quad (3.26)$$

where  $p_{ov,3} > 0$  and  $p_{ov,4} > 0$  are parameters. The sum over neighbours includes effects from bonds that have  $\pi$ -character. The energy penalty for overcoordination is then calculated as

$$E_{over} = \frac{\sum_{j=1}^{nb(i)} p_{ov,1} D_e^\sigma BO_{ij}}{\Delta^{lpc} + Val_i} \Delta^{lpc} \left[ \frac{1}{1 + e^{p_{ov,2} \Delta^{lpc}}} \right] \quad (3.27)$$

where  $p_{ov,1} > 0$  and  $p_{ov,2} < 0$  are parameters and  $D_e^\sigma$  is the same parameter as in equation 3.21. Since the energy penalty of overcoordination is reduced when the energy penalty of lone pair splitting increases, we have two competing effects that we can use to let atoms take different coordination states, which is important for several atoms, e.g. for transition metals and phosphorus.

Undercoordination is usually not a problem, as the energy decreases by the number of bonds formed (by equation 3.21) thus favouring overcoordinated structures. However, the extra free electrons can be stabilised by resonance with neighbouring free electrons, and ReaxFF therefore introduces a stabilisation term for undercoordinated atoms which is given as

$$E_{under} = -p_1^{un} \frac{1 - e^{p_2^{un} \Delta^{lpc}}}{1 + e^{-p_2^{ov} \Delta^{lpc}}} \frac{1}{1 + p_3^{un} e^{p_4^{un} [\sum_{j=1}^{nb(i)} (\Delta_j - \Delta_j^{lp}) (BO_{ij}^\pi + BO_{ij}^{\pi\pi})]}} \quad (3.28)$$

where  $p_1^{un} > 0$ ,  $p_2^{un} > 0$ ,  $p_3^{un} > 0$ ,  $p_4^{un} > 0$  are parameters and  $p_{ov,2}$  is the same parameter as in equation 3.27.

These are the energy terms that ReaxFF applies for bond description, and the important bond order parameter has been demonstrated. The remaining energy terms are given in appendix A together with a list of all the parameters employed by ReaxFF.

### 3.4 Electronegativity Equalisation

ReaxFF employs a Electronegativity Equalisation Method (EEM)[76] (similar to QEq[77]) to calculate the charge of each atom, which later is to be used in the calculation of electrostatic energy. This redistribution of charges gives the force field the ability to let atoms respond to their electric environment, an effect denoted as polarisation. Polarisation effects are only included to a certain degree, as the charges are only distributed over the atomic centers, thus a planar molecule does not have the ability to polarise out of the plane. The

EEM lets the charges of each atom fluctuate by looking at the energy penalty for the transfer of charge to an atom. We approximate the charge transfer energy  $E_{qt}$  of each atom as a Taylor series

$$E_{qt} \approx E_{qt}^0 + \left(\frac{\partial E_{qt}}{\partial q}\right)_0 (q - q^0) + \left(\frac{\partial^2 E_{qt}}{\partial q^2}\right)_0 \frac{(q - q^0)^2}{2} + \dots = E_{qt}^0 + \chi^0 (q - q^0) + \eta^0 (q - q^0)^2 + \dots \quad (3.29)$$

where  $q$  is the charge of the atom and  $\chi_0$  and  $\eta_0$  are called the "electronegativity" and the "chemical hardness" of the atom. We cut the Taylor series after the "hardness" term and get the total electrostatic energy as

$$E_{el} = \sum_i [E_{qt,i}^0 + \chi_i^0 (q_i - q_i^0) + \eta_i^0 (q_i - q_i^0)^2 + \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 |r_i - r_j|}] \quad (3.30)$$

To find the optimal distribution of charges, we attempt to minimise equation 3.30. We also need to keep the total charge of our system constant, and we therefore introduce a lagrange multiplier in the energy equation and get

$$L = \sum_i [E_{qt,i}^0 + \chi_i^0 (q_i - q_i^0) + \eta_i^0 (q_i - q_i^0)^2 + \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 |r_i - r_j|}] - \lambda (\sum_i q_i - Q_{tot}) \quad (3.31)$$

We then minimise equation 3.31 with respect to  $q_i$  and get

$$\chi_i^0 + \eta_i^0 (q_i - q_i^0) + \sum_j \frac{q_j}{4\pi\epsilon_0 |r_i - r_j|} = \lambda = \frac{\partial E_{el}}{\partial q_i} = \chi_i \quad (3.32)$$

where we see that all  $\chi_i$ 's are equal when the energy is at a minimum, thus the name Electronegativity Equalisation Method. This gives us a set of  $N_{at} - 1$  linear independent equations and we know that

$$\sum_i q_i = Q_{tot} \quad (3.33)$$

which gives us a total of  $N_{at}$  linear independent equations, which can be solved by a simple matrix inversion to obtain the charges  $q_i$ . When the number of atoms in the system grows large, the inversion will require a large number of operations (scales as  $\mathcal{O}(N_{at}^3)$ ). As the charges have to be recalculated for every step in the simulation, the EEM will take a significant part of the computational effort.

### 3.5 Simulations

By developing the force field, we have approximated the electronic part of the energy and we have the total energy from equation 2.23 as

$$V_{tot} = E_{kin,n} + \sum_{i<j} \frac{e^2 Z(i)Z(j)}{4\pi\epsilon_0 |r_i - r_j|} + V_{el}(\mathbf{r}^n) \quad (3.34)$$

where  $E_{kin,n}$  is the kinetic energy of the nuclei. The force field usually describes the total potential, thus both the two last terms of equation 3.34, and we have

$$V_{tot} = E_{kin,n} + V_{FF}(\mathbf{r}^n) \quad (3.35)$$

where  $V_{FF}(\mathbf{r}^n)$  is the potential described by the force field. Now we are able to apply Newton's equation (equation 2.2) and let the nuclei propagate over time. To do this we need the gradient of the force field with respect to the nuclear coordinates, but since  $V_{FF}(\mathbf{r}^n)$  is given by analytical functions, the derivatives are

readily accessible.

We obtain the velocities and positions of the nuclei by a numerical integration procedure, such as the Velocity-Verlet[78, 79] routine or the Leap-Frog[80] routine.

## 4 Algorithms

Several algorithms have been implemented in the ReaxFF program package namely

1. Root Mean Square Distance (RMSD) fit
2. Quasi-Newton method for energy minimisation
3. Nudged Elastic Band (NEB) method for calculation of reaction paths

The theoretical background of these methods is displayed in this chapter to give the reader and the potential user of these implemented methods an explanation on what the implemented algorithms actually do. The use of the RMSD-fit method and the NEB-method is demonstrated in the results chapter (Chapter 6).

### 4.1 RMSD fit

One method that has been implemented in the ReaxFF program package is a Root Mean Square Distance (RMSD) fit, that basically takes two molecules and attempts to put the molecules on top of each other by translating and rotating the molecules. This is important for comparison of molecules as translation and rotation can make two similar molecules look quite different. The method of performing the RMSD-fit was given by Kabsch[8] and has been implemented in many programs, however as the proof of the method of Kabsch is a bit tedious, a proof based on Singular Value Decomposition (SVD) is used, based on the proof of Ho[7], but with some further explanations

#### 4.1.1 Minimisation of distance

We want to minimise the distance between corresponding atoms on different molecules by rotation and/or translation. We define the two set of coordinates as

$$y_i = \begin{pmatrix} y_{xi} \\ y_{yi} \\ y_{zi} \end{pmatrix}, x_i = \begin{pmatrix} x_{xi} \\ x_{yi} \\ x_{zi} \end{pmatrix} \quad (4.1)$$

where  $y_i$  is the  $xyz$ -coordinates of atom  $i$  on molecule  $y$ . We define the distance between the two molecules as the sum of the squared atom-atom distances

$$dist = \sum_i [x_i - y_i]^T [x_i - y_i] \quad (4.2)$$

and we want to minimise this distance by applying translation and rotation to one of the molecules. A translation of the molecule will correspond to change all  $xyz$ -coordinates for a molecule by a certain amount, e.g.  $y_i$  after translation can be written

$$y_i^{trans} = y_i + a \quad (4.3)$$

where  $a_i = [a_x a_y a_z]^T$ . The optimal alignment is simply found by aligning the coordinate centers, which can be written

$$\sum_i y_{xi} = \sum_i x_{xi} = 0 \quad (4.4)$$

for all  $x, y, z$ . For simplicity, we align the coordinate centers to the origin. A rotation around the origin can be written as an orthogonal matrix

$$y_i^{rot} = Uy_i \quad (4.5)$$

The rotation matrix is orthogonal, thus  $U^T U = I$  and has determinant = 1. A determinant of  $-1$  will correspond to a rotation and a mirroring. We want to keep the chirality, and not allow any mirroring. We are left with the following minimisation problem

$$\begin{aligned} &\text{minimise: } \sum_i [x_i - Uy_i]^T [x_i - Uy_i] \\ &\text{with respect to: } U \\ &\text{subject to: } U^T U = I \\ &\text{subject to: } \det(U) = 1 \end{aligned} \quad (4.6)$$

#### 4.1.2 Optimal rotation

The optimal rotation is a bit more difficult to find than the optimal translation. We start with equation 4.6 and get.

$$\sum_i [x_i - Uy_i]^T [x_i - Uy_i] = \sum_i [x_i^T x_i - 2x_i^T Uy_i + y_i^T y_i] \quad (4.7)$$

Here, only the term  $2x_i^T Uy_i$  depends on the rotation matrix, and we simplify the minimisation problem to a problem of maximising this term.

$$\text{max: } \sum_i x_i^T Uy_i = \text{Tr}(X^T UY) = \text{Tr}(UYX^T) = \text{Tr}(UR) \quad (4.8)$$

where  $Y = [y_1 y_2 \dots y_n]$ ,  $X = [x_1 x_2 \dots x_n]$  and  $R = YX^T$ . Note that we start with the trace of a  $N_{at}$  by  $N_{at}$  matrix, but after the recasting we only have a 3 by 3 matrix<sup>6</sup>. We decompose R by a singular value decomposition as

$$R = VSW^T \quad (4.9)$$

,where  $S$  is a diagonal matrix with positive entries, and  $V$  and  $W^T$  are orthogonal matrices such that

$$VV^T = I = WW^T \quad (4.10)$$

By inserting the SVD decomposition of R into equation 4.8 we obtain

$$\text{max: } \text{Tr}(UVSW^T) \quad (4.11)$$

as all these matrices are square, we can cycle them around and obtain

$$\text{max: } \text{Tr}(SW^T UV) = \text{Tr}(ST) \quad (4.12)$$

---

<sup>6</sup>Rules for recasting and juggling with traces of matrices can be found in Wikipedia at <http://en.wikipedia.org/wiki/Trace>, but not to be mixed up with trace used in the science of "deconstruction". Quoting Wikipedia: "in deconstruction, a trace is the self-effacing signified that a sign differs and defers from". Of course....

The matrix  $T$  is a product of orthogonal matrices ( $U$  has to be orthogonal to be a rotation matrix), and is itself an orthogonal matrix. The matrix  $S$  has only diagonal elements, thus we obtain

$$\max : s_{11}t_{11} + s_{22}t_{22} + s_{33}t_{33} \quad (4.13)$$

as our target function, where  $t_{ij}$  and  $s_{ij}$  are elements in the matrices  $S$  and  $T$  respectively. The restriction on the optimisation is that the  $T$ -matrix is orthogonal, such that

$$\begin{aligned} t_{11}^2 + t_{21}^2 + t_{31}^2 &= 1 \\ t_{12}^2 + t_{22}^2 + t_{32}^2 &= 1 \\ t_{13}^2 + t_{23}^2 + t_{33}^2 &= 1 \\ t_{11}t_{12} + t_{21}t_{22} + t_{31}t_{32} &= 0 \\ t_{11}t_{13} + t_{21}t_{23} + t_{31}t_{33} &= 0 \\ t_{12}t_{13} + t_{22}t_{23} + t_{32}t_{33} &= 0 \end{aligned} \quad (4.14)$$

The  $T$ -matrix consists of 9 elements, and we have 6 restricting equations (equations 4.14). By a closer look at these equations we see that we can pick 3 elements from different rows and columns and let them vary freely between  $-1$  and  $1$ . We choose the diagonal elements  $t_{11}$ ,  $t_{22}$  and  $t_{33}$  to be our free variables and get the optimisation problem

$$\begin{aligned} \max : s_{11}t_{11} + s_{22}t_{22} + s_{33}t_{33} \\ \text{s.t.} : -1 \leq t_{11}, t_{22}, t_{33} \leq 1 \end{aligned} \quad (4.15)$$

This problem is a linear optimisation problem and we see that all the extremals must be on the “edge”, where the  $t_{ii}$ ’s meet their restricting limit (except when some of the  $s_{ii}$ ’s are identical). Thus the  $t_{ii}$ ’s can only obtain the values  $-1$  and  $1$ . Since the  $s_{ii}$ ’s are all non-negative, the solution of equation 4.15 will give all  $t_{ii} = 1$ , and  $T$  will be the identity matrix.

We can then solve for  $U$

$$T = I = W^T U V \quad (4.16)$$

$$W W^T U V V^T = W I V^T = W V^T = U \quad (4.17)$$

We have until now only put the restriction on  $U$  that it should be orthogonal. Both rotation matrices and mirror matrices can be written as orthogonal matrices, thus  $U$  obtained by equation 4.17 can also contain mirror planes. We want to restrict  $U$  to be only proper rotations, and this will be the case if  $\det(U) = 1$ . If on the other hand,  $\det(U) = -1$  we need to look for another solution. The second largest value equation 4.15 can obtain will come if the smallest of  $s_{ii}t_{ii}$  changes sign (remember that  $t_{ii} = -1$  or  $1$ ). This will also change the sign of  $\det(U)$  and we will have a proper rotation. The optimal  $U$  in such a case will then be

$$W W^T U V V^T = W I^* V^T = U \quad (4.18)$$

where  $I^*$  is the identity matrix where the diagonal element corresponding to the lowest  $s_{ii}$ -value is  $-1$ . Thus by creating the  $R$  matrix (equation 4.8) and performing an SVD, we can easily obtain the optimal rotation matrix by using equation 4.17 or 4.18 and a RMSD fit procedure based on SVD is implemented in the ReaxFF program package.

## 4.2 Removal of translation and rotation

Optimisation methods can be performed in a variety of coordinate systems. Cartesian coordinates (xyz) is an intuitive set of coordinates, since they form an orthogonal set and are easy to work with. We are also able to define the atom coordinates as a number of bond lengths, bond angles and dihedral angles which are referred to as internal coordinates. The methods for energy minimisation implemented in the ReaxFF package are all based on cartesian coordinates. A molecule is defined by 3 coordinates per atom, thus being in a  $3N_{at}$ -dimensional space. The chemical properties of a molecule do not change by translation and rotation, thus we actually have 6 cartesian coordinates more than we need. The minimisation routines are still able to perform even if we have 6 coordinates that should be unnecessary. However in the Nudged Elastic Band method that will be explained in Chapter 4.4 we use a distance metric, which will need us to prevent translation and rotation of the molecule. A method that has been used to remove translation and rotation is to fix one atom, then restrict a second atom to move in a plane, and restrict a third atom to move on a line thus restricting a total of 6 coordinates[81, 10]. Another approach is to remove the forces that would lead to translation and rotation by a projection. The problem of finding which forces to remove to prevent translation is trivial and shown in the following chapter. A proof of how to find the directions where the forces should be zero to prevent rotation is given in Chapter 4.2.2. As far as the author knows, this formulation of the proof has not been proposed earlier.

### 4.2.1 Removal of translation

We assume that we have calculated the gradient  $g$  of the energy for a set of atoms, and make a small step in the opposite direction of the gradient. We want to change the gradient such that the coordinate center is constant during the step. The step can be written

$$c_{i+1}^n = c_i^n - kg_i^n \quad (4.19)$$

for each atom  $n$ . We want to modify  $g_i$  such that

$$\sum_n c_{i+1}^n = \vec{0} = \sum_n c_i^n - k \sum_n g_i^n = -k \sum_n g_i^n \quad (4.20)$$

thus the sum of all forces has to be 0 for all three coordinates. The correction for the x-coordinate is shown below. We define

$$g_i^x = g_i^{x0} + \Delta g_i^x \quad (4.21)$$

To satisfy the criteria in equation 4.20 we set

$$\Delta g_i^x = -\frac{\sum g_i^{x0}}{N} \quad (4.22)$$

and perform the correction for the x,y and z coordinates. We are now restricting 3 coordinates of our set of atoms, and we have now reduced our space to  $3N_{at} - 3$  dimensions.

### 4.2.2 Removal of rotation

We approach the problem in the same manner as for the removal of translation forces, by assuming a step of the form in equation 4.19. To remove rotation from the gradient we request that the optimal rotation matrix  $U$  arising from Chapter 4.1.2 by comparing the coordinates before and after the step equals the identity matrix, thus no rotation has been performed. We request that

$$U = WV^T = I \quad (4.23)$$

Since the  $W$  and  $V$  matrices are orthogonal, this gives

$$W = V \quad (4.24)$$

and  $R$  (equation 4.8) can be written as

$$R = WSW^T \quad (4.25)$$

and will be a symmetric matrix. Thus the criteria we want to fulfill is that the  $R$ -matrix after the step is a symmetric matrix. The  $R$ -matrix as defined in equation 4.8 becomes

$$R = C_{i+1}C_i^T \quad (4.26)$$

where  $C_{i+1} = [c_{i+1}^1 c_{i+1}^2 \dots c_{i+1}^N]$  and  $C_i = [c_i^1 c_i^2 \dots c_i^N]$ .  $C_{i+1}$  can be rewritten as

$$C_{i+1} = C_i - kG_i \quad (4.27)$$

where  $G_i = [g_i^1 g_i^2 \dots g_i^N]$  by assuming a step in the opposite direction of the gradient. Equation 4.26 then gives

$$R = C_i C_i^T - kG_i C_i^T \quad (4.28)$$

The first term  $C_i C_i^T$  is symmetric, thus the last term must also be symmetric. From now on, the indices  $i$  will be removed to ease the reading. Our restriction on  $G$  will be that it fulfills

$$GC^T = CG^T \quad (4.29)$$

This leads to the following set of restrictions

$$\begin{aligned} g_x^T c_y &= g_y^T c_x \\ g_x^T c_z &= g_z^T c_x \\ g_y^T c_z &= g_z^T c_y \end{aligned} \quad (4.30)$$

where  $g_x = [g_x^1 g_x^2 \dots g_x^N]^T$ ,  $c_x = [c_x^1 c_x^2 \dots c_x^N]^T$  and analogous for  $y$  and  $z$ . Equations 4.30 can be rewritten as

$$\begin{bmatrix} c_y^T & -c_x^T & 0 \\ c_z^T & 0 & -c_x^T \\ 0 & c_z^T & -c_y^T \end{bmatrix} \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} = 0 = Ag \quad (4.31)$$

Thus,  $g \in \text{Null}(A)$ . To make  $g$  lie in the Null space of  $A$ , we project away the part of the gradient that does not lie in  $\text{Null}(A)$ , which will be equivalent to project away the part of the gradient that lies in the Row space of  $A$ . In other words

$$g = g^0 - g_{projRow(A)}^0 \quad (4.32)$$

where  $g^0$  is the uncorrected gradient, and  $g_{projRow(A)}^0$  is the projection of  $g^0$  onto the Row space of  $A$ . The matrix  $A$  has three rows, and the projection can be performed by orthogonalizing the row vectors of  $A$ , and then projecting  $g^0$  onto the orthogonal basis. By this projection, we have removed 3 coordinates more and a total of 6 coordinates.

### 4.3 Quasi-Newton method

A Quasi-Newton method has been implemented in the ReaxFF package for energy minimisation, because of its computational advantages over other gradient based minimisation methods (e.g. steepest descent and conjugate gradient). A Newton-method for optimisation is based on the assumption of a quadratic approximation of the minima. We approximate the function to minimise by a Taylor expansion as

$$V(x) \approx V(x_0) + \nabla V(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 V(x_0)(x - x_0) + \dots \quad (4.33)$$

where  $x$  and  $\nabla V(x_0)$  have the form of vectors, and  $\nabla^2 V(x_0)$  has the form of a matrix, and is denoted "the Hessian matrix" or only "the Hessian". We differentiate  $V(x)$  and get

$$\nabla V(x) \approx \nabla V(x_0) + \nabla^2 V(x_0)(x - x_0) + \dots \quad (4.34)$$

We are searching for the minima, and by neglecting higher order terms and requesting  $\nabla V(x) = 0$  we obtain

$$x = x_0 - (\nabla^2 V(x_0))^{-1} \nabla V(x_0) = x_0 - H_0^{-1} g_0 \quad (4.35)$$

where  $g_0 = \nabla V(x_0)$  and  $H_0 = \nabla^2 V(x_0)$ . This method of optimisation is called the Newton method. If we know the Hessian and the potential really is quadratic, we will hit the minima in one step. However, the potential is normally not quadratic, and the calculation of the Hessian is expensive. Thus we approximate the Hessian for every step by using the gradient from the last step. This estimating procedure of the Hessian has given the name "Quasi-Newton optimisation". The updating procedure will hopefully improve the Hessian after every step, and give a more and more accurate optimisation step. The updating procedure of the Hessian can be done in many ways, and two famous formulas for the updating procedure are presented here. The most common update formula for minimisation problems is the BFGS[82, 83, 84, 85] update formula, which is given as

$$H_{k+1} = H_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} \quad (4.36)$$

where  $y_k = g_{k+1} - g_k$  and  $s_k = x_{k+1} - x_k$ . Another formula which is commonly used for transition state searches is the Powell [86] update formula which is given as

$$H_{k+1} = H_k + \frac{1}{s_k^T s_k} [m_k s_k^T + s_k m_k^T - \frac{m_k^T s_k s_k s_k^T}{s_k^T s_k}] \quad (4.37)$$

where  $m_k = y_k - H_k s_k$ . Both these update formulas are based on the condition that

$$H_{k+1} s_k = y_k \quad (4.38)$$

thus the new Hessian is updated such that it describes the change in gradient by the last step performed. The BFGS update formula is constructed such that if  $y_k^T s_k > 0$ , then the update will not remove any positive definiteness of the Hessian. A matrix is positive definite if

$$x^T H x \geq 0 \forall x \quad (4.39)$$

Since the Hessian is symmetric, this will mean that all the eigenvalues of  $H$  are positive. Note that in the derivation of the Newton formula (equation 4.35), we only requested the gradient to be 0, which will be the case for all extremals including the saddle points, and a Newton step can lead us to any of those points. Normally we are interested in finding the minima, and thus minimising along all eigenvectors of  $H$ . By using the newton step (equation 4.35) this will happen when all eigenvalues of  $H$  are positive, thus  $H$  is positive definite. The BFGS-update formula ensures that the Hessian is always positive definite and therefore the

newton step calculated will always be in a descent direction.

The Powell update, does not ensure a positive definiteness, which will be useful in finding transition states which are saddle points and not minimas. It also allows the eigenvalues of  $H$  to change rapidly, which is a useful property for transition state searches like the eigenvector following method[87, 88, 89].

### 4.3.1 Line search

To compensate for the non-quadratic form of the potential and to increase stability, a Quasi-Newton method is mostly performed with a line search in the direction  $-H_0^{-1}g_0$ . A common way to do a line search is to do a step, then fit the energy and gradients calculated for the step together with the old energy and gradient to a polynomial, and then use the minima on this polynomial as the improved step. Another common method is to use a trust region procedure [90], where the calculated energy change is compared to the exact energy change of a step. If the deviations are too large, the step is rejected and a smaller step is attempted. The final approach is to rely on the Newton step to scale itself when the Hessian is accurate enough, thus only restricting the step to not exceed some maximal length, and hope that the Newton step finally manage to hit the minima.

A Quasi-Newton method for minimisation of structures based on the BFGS-update formula (equation 4.36) has been implemented in the ReaxFF program package.

## 4.4 Nudged elastic band

A common method for finding transition states is the Nudged elastic band (NEB) method [9, 10]. This method approximates the Minimum Energy Path (MEP) between structures. The MEP will be the most probable reaction coordinate between the two structures, as it demands the lowest activation energy. If there really is an energy barrier between the two structures, the MEP will cross the barrier at the lowest possible energy, thus give the transition state with the lowest possible energy<sup>7</sup>.

The NEB method demands two input structures, one for the reactant structure and one for the product structure. The NEB method starts by creating images as a linear interpolation between the two provided structures. We define the coordinates of the images as  $[C_0, C_1, \dots, C_N]$  where  $C_0$  and  $C_N$  are the two input images. The  $[C_1, \dots, C_{N-1}]$  images will most probable not be on the MEP, and we want to move them in the direction of the MEP. The gradient of each image can be calculated, but by letting the images follow the gradient, all the images will go in the direction of the minimas. We therefore only let the gradient work perpendicular to the MEP. We also want to keep the images distributed over the MEP and we therefore create a spring force to work along the MEP. We have

$$F_i = F_i^{\parallel} + F_i^{\perp} \quad (4.40)$$

where  $F_i^{\parallel}$  is the spring force that acts along the MEP and  $F_i^{\perp}$  is the part of the gradient orthogonal to the MEP. However, we don't have a MEP yet, and we don't know in which direction it goes. We therefore need to approximate this path. A natural choice of the tangent would be a linear combination of the nearby images as

$$\tau_i = \frac{C_i - C_{i-1}}{|C_i - C_{i-1}|} + \frac{C_{i+1} - C_i}{|C_{i+1} - C_i|} \quad (4.41)$$

---

<sup>7</sup>Imagine you have to cross the alps by foot and try to follow a straight line. You will most definately walk in the valleys instead, thus following the MEP. But without a map it is not always obvious where the MEP is, and you are walking in three simple dimensions, not in  $3 N_{at} - 6$  dimensions. Imagine having to find the MEP in  $3 N_a - 6$  dimesions.... However, some deviations from the MEP are highly recommended if you are hiking in the alps

However, it has been shown that the NEB-optimisation is more stable by making the tangent based on the energy of the images[91] as

$$\tau_i = \begin{cases} \tau_i^+ & \text{if } V_{i+1} > V_i > V_{i-1} \\ \tau_i^- & \text{if } V_{i+1} < V_i < V_{i-1} \end{cases} \quad (4.42)$$

where  $\tau_i^+ = C_{i+1} - C_i$  and  $\tau_i^- = C_i - C_{i-1}$ . In other words we approximate the tangent by only looking at the neighbour with the highest energy. Two cases are not described by equation 4.42, if both or no neighbours are higher in energy (if  $V_{i-1} > V_i < V_{i+1}$  or  $V_{i-1} < V_i > V_{i+1}$ ). For such cases we define the tangent as

$$\tau_i = \begin{cases} \tau_i^+ \Delta V_i^{max} + \tau_i^- \Delta V_i^{min} & \text{if } V_{i+1} > V_{i-1} \\ \tau_i^+ \Delta V_i^{min} + \tau_i^- \Delta V_i^{max} & \text{if } V_{i+1} < V_{i-1} \end{cases} \quad (4.43)$$

where  $\Delta V_i^{min} = \min(|V_{i+1} - V_i|, |V_i - V_{i-1}|)$  and  $\Delta V_i^{max} = \max(|V_{i+1} - V_i|, |V_i - V_{i-1}|)$ . We can then form the spring force  $F_i^{\parallel}$  as

$$F_i^{\parallel} = k(|C_{i+1} - C_i| - |C_i - C_{i-1}|)\tau_i^N \quad (4.44)$$

where  $k$  is a constant and superscript  $N$  denotes normalized. This definition of the spring force can be understood as minimising a quadratic potential

$$V_{spring} = \sum_i \frac{1}{2} |C_{i+1} - C_i| \quad (4.45)$$

which will have its minimum when the images are separated by the same distance. It is also natural to define our MEP as a function of the distance metric  $|C_{i+1} - C_i|$ . This potential will also have a minimum for a shortest possible path, but since the forces from the spring potential only work along the MEP, this will have no effect. We let the gradient work perpendicular to the MEP and we get

$$F_i^{\perp} = -(\nabla V - \nabla V \cdot \tau_i^N) \quad (4.46)$$

We are now able to construct the total forces on each image from equation 4.40. By letting our system evolve due to these forces, all the images should move closer to and hopefully end on the MEP.

#### 4.4.1 NEB-BFGS

Since now all the images are coupled to each other through  $F_i^{\parallel}$ , we now have  $3N_{at}(N_{im} - 2)$  variables in our system. Quasi-newton methods, especially coupled to the BFGS update (equation 4.36) have been shown to be effective methods of minimisation of potentials. The NEB-method is not coupled to an overall potential function, but a BFGS quasi-newton method has been successfully employed on the NEB-method, though it is reported to have some convergence problems[81].

The spring potential (equation 4.45) is defined as a function of the distance metric  $|\cdot|$ , and chemically identical structures can be separated by a distance if they are translated or rotated compared to each other. We therefore want to minimise the initial distance between two structures, and the input structures are therefore aligned to each other by a RMSD-fit (Chapter 4.1). Also the translation and rotation is removed from the forces by the method described in Chapter 4.2 before it is applied to the images and the Hessian updating procedure. An alternative approach could be to perform a RMSD-fit after every step to prevent the molecule from rotating and translating, instead of removing the forces that lead to the rotation and translation. This will however create an inconsistency between the performed step and the gradient used for optimisation, thus conflicting with the Hessian updating procedure.

A NEB-method based on the NEB-BFGS combination has been implemented in the ReaxFF program package.

## 5 Computational details

### 5.1 Quantum chemistry calculations

Quantum chemistry calculations have been performed with the B3LYP DFT functional (Chapter 2.50) and the 6-311++G(d,p) basis set (Chapter 2.4.1). Both Gaussian 03[47], Gaussian 98[92] and Jaguar[48] has been used during this project. The Gaussian calculations have been performed with a ultrafine[47] grid for numerical integration, while the Jaguar calculations have been performed with the default[48] grid. All structures have been minimised, and for reaction paths, 1 coordinate has been kept fixed (e.g. bond length, dihedral angle) while all others have been optimised. For radicals, the spin multiplicity has been set to 2, except for the methanol-oxygen system, where the multiplicity has been set to 3, as the ground state of oxygen is a triplet and the end product of the reaction is two radicals. For all the open-shell systems restricted open-shell B3LYP has been used. Mulliken[11] charges have been calculated for some of the geometry optimised structures with a 6-31G(d,p) basis set. The DFT energies are not corrected for BSSE.

### 5.2 Parametrisation

The parametrisation to ReaxFF has been performed by Adri van Duin at the Materials and Process Simulation Center, Caltech, Los Angeles. The carbon and hydrogen parameters have been left unchanged from previous parametrisations[6], thus only parameters involving oxygen have been parametrised. This includes the C-O and H-O bonds, valence angles involving oxygen (e.g. C-C-O, C-O-C, C-O-H) and torsion angles involving oxygen (e.g. C-O-C-H, C-C-O-H, etc.). The parametrisation has been performed against data obtained by DFT calculations, and has been performed by letting ReaxFF relax due to the same constraints (such as fixed bond lengths, angles etc.) as the DFT calculations. This leads to some deviations in the structure of the molecules obtained by DFT and the structure that ReaxFF parametrise after. The parametrisation procedure demands the assignment of weights to the different data in the training set, then the force field optimisation procedure attempts to minimise the errors by varying one parameter at the time. The training set consists of a large number of data involving

1. Atomic charges  
Mulliken charges of  $\beta$ -D-glucopyranose, cellobiose and  $\text{H}_2\text{O}$ .
2. Geometry parameters  
Selected bond distances, valence angles and dihedral angles of  $\beta$ -D-glucopyranose and the bond distance of the linkage between glucose monomers of cellobiose.
3. Energies  
Energies of a large number of structures (1555 minimised or constrained minimised structures) with corresponding DFT energies. These include the energies of rotations, dissociations and oxidations of small carbon-hydrogen-oxygen molecules, the energies of small carbon-hydrogen-oxygen radicals, energies of  $\text{H}_2\text{O}/\text{H}_3\text{O}^+$  clusters, proton transfer between water molecules and many more. A set of carbohydrate specific reactions and structures were also included in the training set.

The number of structures included in the training set is so large that a complete presentation of the parametrised structures and the ReaxFF fit will not be possible. Only the fit to the specific carbohydrate structures will be presented in Chapter 6, together with model systems for oxidation and protonation reactions.

### 5.3 Test of force field

For the cellulose unit cell optimisation by ReaxFF the cellulose crystal structure I<sub>α</sub>[93] has been taken and the energy has been minimised by expanding/compressing the unit cell lengths until a minimum has been found.

The periodic box prepared for molecular dynamics simulations contains 4 cellulose strains, each with the structural formula C<sub>24</sub>H<sub>40</sub>O<sub>20</sub>, together with water molecules and a varying number of H<sub>3</sub>O<sup>+</sup> and Cl<sup>-</sup> molecules, keeping the system neutral. The total number of H<sub>2</sub>O and H<sub>3</sub>O<sup>+</sup> molecules is 101. The total volume of the periodic box is about 7000 Å<sup>3</sup>, and the water phase is about 60 % of the total volume, thus giving a density of about 0.7 kg/dm<sup>3</sup> for the water phase. Initial structures for molecular dynamics simulations have been energy minimised and equilibrated at 300 K for 2.5 ps. The structures have been heated at a rate of 0.4 K/fs, and then run at the designated temperature for a total of 25 ps. All simulations have been performed with a time step of 0.25 fs and a Berendsen thermostat[94] has been employed to adjust the temperature during simulations.

## 6 Results and discussion

The results of the parametrisation will be presented here, though only the carbohydrate related structures and model systems for oxidation and protonation. There are a couple of reasons to this, firstly the training set for the parametrisation consists of a substantial number of structures and a complete discussion could fill a report by itself. Secondly only a part of the training set has been prepared as a part of this project, and thirdly this part of the training set consist of the most relevant structures for carbohydrate comparison, and will therefore give the most knowledge on the performance of the force field when it comes to carbohydrates.

The second part of the results gives a demonstration on which kind of applications a working force field has. Some simple simulations and calculation of reaction paths have been performed, and defects of the force field have been identified from the results of these calculations.

### 6.1 Parametrisation and comparison to DFT-data

The carbohydrate specific DFT energies used in the training set for parametrisation will here be compared to the corresponding ReaxFF fit. Also model systems for the oxidation and protonation reaction barriers are presented. The idea is to use a model system to get the reaction barriers right, then only compare the relative stabilities of specific structures. The model systems are chosen as carbohydrate-like structures, but with a smaller number of atoms to reduce the time consumption of the DFT-calculations. Even if the DFT energies have been used for parametrisation, it is useful to look at how good ReaxFF fits to the training set. Since the training set itself contains a large number of structures (1555) it is expected that ReaxFF will fit to general features in the training set and not only to the specific features. However, the fit will probably be slightly worse for non-parametrised structures. Some supplementary DFT calculations have been performed which have not been used in the parametrisation which are displayed in Chapter 6.1.3. All other reaction barriers and structures have been included in the training set.

The systems studied are

1. Model system for oxidation

The attack of oxygen (O<sub>2</sub>) on methanol has been chosen as a model system for the oxidation reaction barrier. Also the stabilities of relevant carbohydrate radicals are presented.

## 2. Model system for protonation

The protonation of tetrahydropyran has been chosen as a model system for the protonation reaction barrier. Also the stabilities of relevant protonated carbohydrate structures are presented

## 3. Dissociation of protonated ether oxygen

After a protonation of an ether oxygen, a probable reaction step is the dissociation of one of the oxygen bonds. Two such reactions are presented.

## 4. Carbohydrate conformations

Rotation barriers, the reaction barrier of a ring inversion of the glucose-ring and the stabilities of a set of glucose isomers and conformers are presented.

Different glucose conformers have been reported by Appell et al.[67] and the minimum structure reported for  $\beta$ -D-glucopyranose has been used as a reference structure (denoted  $\beta$ -gt- $^4C_1$  in the paper by Appell et al.) and is shown in Figure 4.

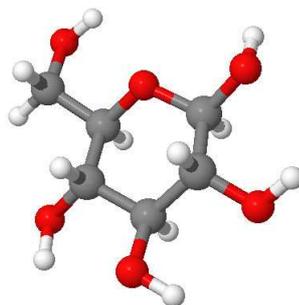


Figure 4: Minimum energy structure of  $\beta$ -D-glucopyranose as reported by Appell et al.[67], and used as a reference for glucose isomers and conformers, and for reactions involving glucose in this project

It is to be emphasised that the reaction barriers presented have been obtained by keeping the same constrained coordinate in the DFT and ReaxFF calculations, then letting all other coordinates relax as explained in Chapter 5.

### 6.1.1 Oxidation

As a model system for oxidation, the reaction between methanol and oxygen has been chosen. Two reactions have been considered for this system,

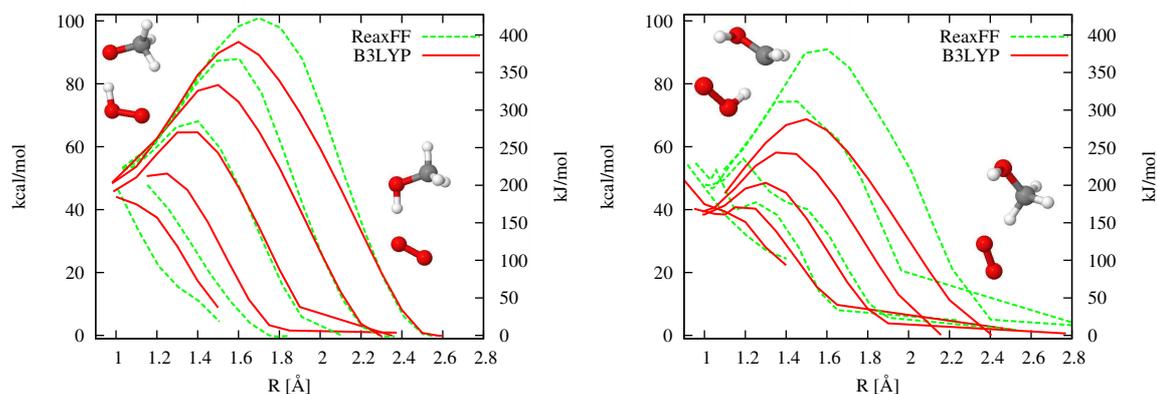


and



The energies of the reactions have been calculated by keeping a fixed distance between  $O - O$  and  $C - O$  respectively, then allowed the hydrogen to move in between. This procedure has been performed for different fixed distances, and the results are shown in Figures 5(a) and 5(b). There are two reasons for keeping these distances fixed, first it is easier to keep these structures during the parametrisation procedure and secondly during a simulation atoms are not fixed and the reaction can also occur when the distance between

the reacting molecules is not optimal. Thus the energy barrier for variable distances between the reacting molecules is also an important parameter.



(a) Energy barrier for fixed O-O distances and varying the distance between the oxygen of  $O_2$  and the hydrogen of the O-H on methanol. The fixed O-O distances are 2.5, 2.75, 3.0, 3.25 and 3.5 Å, where a lower energy barrier corresponds to a shorter fixed distance. (b) Energy barrier for fixed C-O distances and varying the distance between the oxygen of  $O_2$  and the hydrogen of the C-H on methanol. The fixed C-O distances are 2.5, 2.75, 3.0, 3.25 and 3.5 Å, where a lower energy barrier corresponds to a shorter fixed distance.

Figure 5: Energy barriers for oxidation of methanol

There are clearly differences between the fit for reaction 6.1 and 6.2. Reaction 6.1 shows a good fit for most fixed O-O distances. It also predicts the relative stability of reactants and products accurately. This is not the case for reaction 6.2. Here ReaxFF predicts a too high barrier for C-O distances of 3.25 and 3.5 Å. However the energy barrier for the shorter distances fit well to the DFT calculations. ReaxFF also predicts an energy difference of about 50 kcal/mol between reactants and products of reaction 6.2, while the DFT calculations give an energy difference of about 40 kcal/mol. The DFT calculations predict the carbon radical to be slightly more stable than the oxygen radical. A separate calculation of the radical energies shows that DFT predicts the energy of the oxygen radical to be 6.3 kcal/mol higher than the carbon radical while ReaxFF gives the difference to be only about 1.5 kcal/mol, thus ReaxFF overstabilises the oxygen radical. ReaxFF also seems to predict that the oxidation barriers for reaction 6.1 and 6.2 are more similar than the DFT calculations do.

To force the parametrisation to get the stabilities of relevant carbohydrate radicals right, energies of glucose radicals has been calculated and compared. The results of this comparison are plotted in Figure 6. Here we also see that the carbon radicals are more stable than the oxygen radicals, an effect also ReaxFF includes but with a smaller energy difference, as also seen for the methanol radicals. However, ReaxFF seems unable to describe the different stabilities within the groups of oxygen and carbon radicals. The only effect that could possibly be distinguished is that ReaxFF predicts that the O1 radical is the most stable of the oxygen radicals, which is an effect described by the DFT calculations. ReaxFF also seems to overstabilise the carbon radicals vicinal to the ring oxygen (C1 and C5), probably due to a resonance effect between the radical and the ring oxygen.

ReaxFF seems to be able to describe radical reactions well, it shows the correct trends and reasonable energy barriers and stabilities. However there are some discrepancies, e.g. it is not able to predict which of the

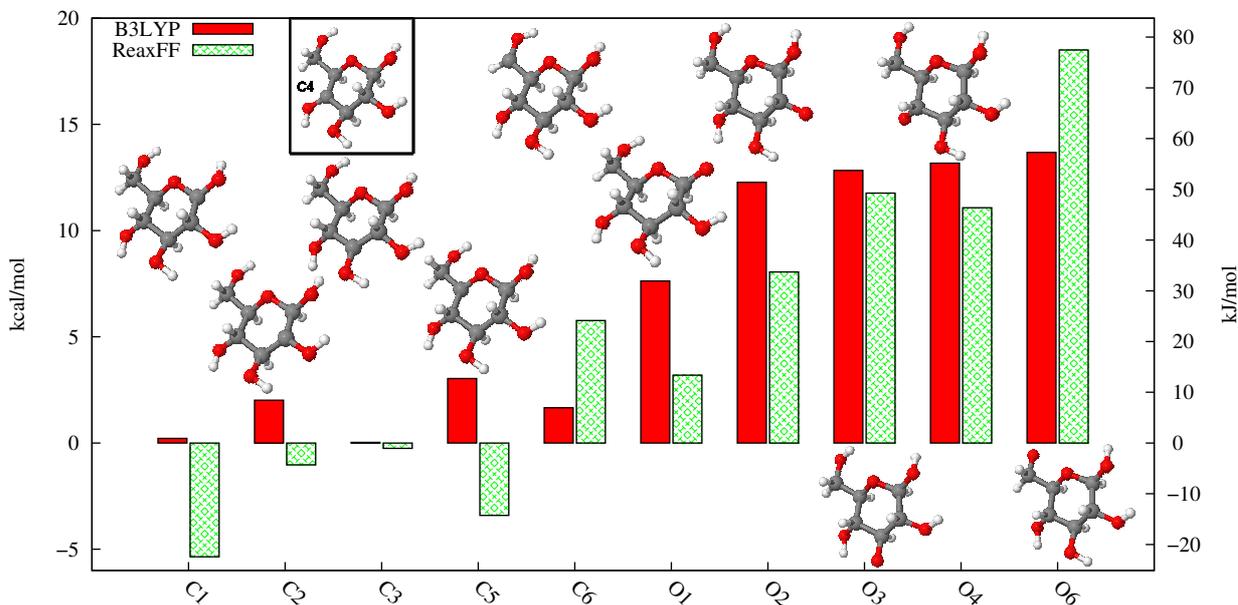


Figure 6: Energy of some isomers of glucose radicals. All structures are similar to the minimum structure of  $\beta$ -D-glucopyranose (Figure 4), except for the removal of a hydrogen.  $CX$  signifies that a hydrogen has been removed from carbon  $X$ , and analogous for  $OX$ . All structures displayed are isomers, and have been compared to the energy of the  $C4$  structure displayed in the figure.

glucose carbon/oxygen radicals that is the most stable. It is therefore expected that ReaxFF can describe radical reactions, though not display the correct ratio between which radicals that are formed which implies that the ratio between different reaction paths also will not be quantitatively correct relative to the DFT calculations.

### 6.1.2 Protonation

The second model system, the protonation of tetrahydropyran ( $C_5H_{10}O$ ), is chosen to describe the protonation of and ether oxygen.

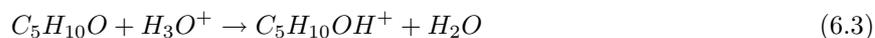
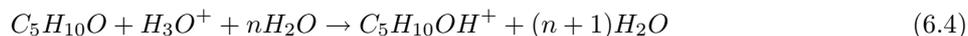


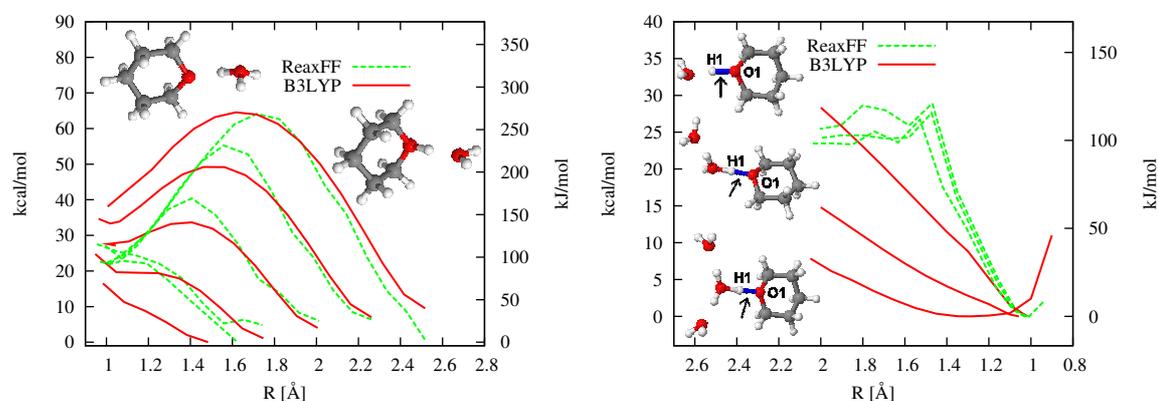
Figure 7(a) shows the energy barrier of transferring a proton from tetrahydropyran to water for fixed O-O distances. The barrier of the reaction is modelled well by ReaxFF. The major deviation from the DFT calculations is the stability of the  $H_3O^+$  ion, as seen by the different energies of the unprotonated structures in Figure 7(a). As the DFT calculations show, this is greatly dependent on the O-O distance, where  $H_3O^+$  is more stable for shorter O-O distances. This is a result of the fact that the positively charged  $H_3O^+$  can share its positive charge with the oxygen of tetrahydropyran, an effect that ReaxFF does not seem to include. Similarly if water is present nearby it is believed that these also can stabilise the positively charged  $H_3O^+$ , and shift the reaction energy barrier. To test this effect some additional quantum chemistry calculations were performed.

### 6.1.3 Additional protonation reactions

To check the effect that nearby water molecules have on the protonation reaction, additional DFT calculations have been performed. These were not included in the parametrisation of the force field. The first set of additional DFT calculations were performed on the same reaction as equation 6.3



by varying the number of non-reacting water molecules from 0 to 2. Figure 7(b) shows reaction 6.4 with 0, 1 and 2 additional water molecules present, and this has a great effect on the DFT results. The energy of  $H_3O^+$  is greatly reduced, and also the minimum bond distance changes radically when 2 water molecules are present. ReaxFF does not show such a stabilising effect of nearby water molecules, an effect that should be attempted to include in further parametrisations.



(a) Energy as a function of the distance between oxygen on water and the proton, where no additional water molecules are present. The O-O distance has been kept fixed as 2.5, 2.75, 3.0, 3.25 and 3.5 Å. A lower energy barrier corresponds to a shorter fixed distance. (b) Energy as a function of the distance between oxygen of tetrahydropyran and the proton, where the number of additional water molecules has been varied from 0 to 2. A lower energy barrier corresponds to a higher number of water molecules present. The O-O distance is allowed to vary freely

Figure 7: Energy barrier of protonation of tetrahydropyran

The hydroxy groups on glucose are expected to behave in the same manner as water molecules, thus shifting the energy barrier of protonation. The second set of additional DFT calculations have been performed on the protonation of the ring oxygen of  $\beta$ -D-glucopyranose, due to the following reaction



The energy of this reaction as a function of the distance between the ether oxygen and the proton is plotted in Figure 8. Here we see that the reaction energy given by DFT is in the opposite direction of the energy predicted by ReaxFF. This result is also believed to be the result of a stabilisation of the nearby water molecule and the hydroxy groups on glucose, completely turning around the reaction profile. This protonation reaction is also advised to include in further parametrisations to include the stabilisation effect that water shows.

For the protonation reactions, ReaxFF describes the reaction energy barriers well, but fails to describe the energy of  $H_3O^+$  properly, as the stabilisation effect of nearby water molecules is not included. The training set did include  $H_2O/H_3O^+$  clusters, but the comparisons between the clusters were performed by comparing

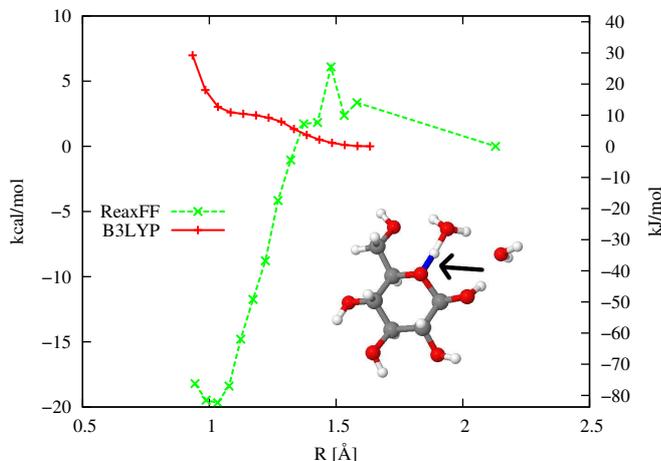


Figure 8: Protonation energy barrier. Energy as a function of distance between ring oxygen in glucose and proton. One water molecule has been kept nearby the reaction site during the calculations.

energies of two separate clusters. E.g. if we have a charged atom  $A^+$  and want to compare the stability of the molecule in vacuum and in a cluster of molecules  $nB$ , the comparison was simply performed by calculating

$$E_{stab} = E(A^+_{alone}) + E(nB_{alone}) - E(A^+/nB) \quad (6.6)$$

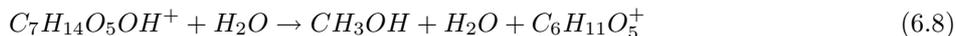
where  $A_{alone}$  and  $nB_{alone}$  were in separate computational boxes, and compared to a computational box where  $A$  and  $nB$  were interacting. Since the system is charged, this requires us to push the whole positive charge onto  $A$  when it is in a separate computational box. In a simulation where the dissociation occurs by separating  $A^+$  and  $nB$  by a large distance, it is not given that the EEM enables to push the whole positive charge onto molecule  $A$ . Thus, to connect the parametrisation with situations relevant for simulations, we should use the comparison

$$E_{stab} = E(A^+/d_{large}/nB_{alone}) - E(A^+/nB) \quad (6.7)$$

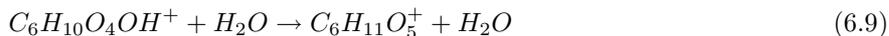
where the energies of a system where  $A$  and  $nB$  are separated by a large distance is compared to the system where  $A$  and  $nB$  interact. This should be modified for the training set for further parametrisations.

#### 6.1.4 Dissociation of ether oxygen

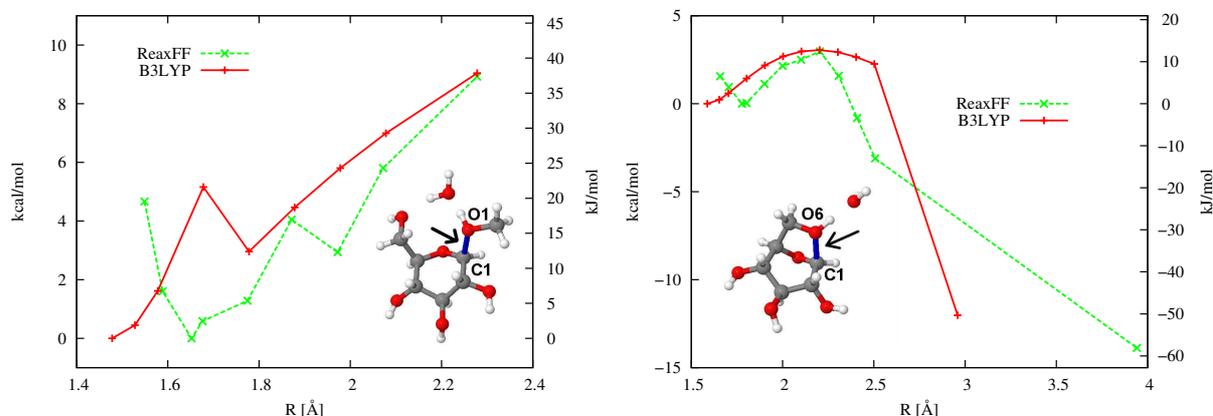
After a protonation of an ether oxygen, a possible reaction is the dissociation of one of the former bonds to oxygen. Two such dissociations have been modelled and used for the parametrisation, firstly the dissociation of the ether linkage of methyl- $\beta$ -D-glucopyranoside giving the reaction



and secondly the dissociation of a protonated structure of levoglucosan (1,6-anhydro- $\beta$ -D-glucopyranose) with the isomerisation reaction



The energy of these dissociation reactions are plotted in Figure 9(a) and 9(b) respectively. The energy barriers of reactions 6.8 and 6.9 are well modelled by ReaxFF. ReaxFF predicts both the energy barrier and the total stabilisation well. The major difference between the DFT calculations and ReaxFF calculations is



(a) Energy barrier for the opening of the ether linkage of methyl- $\beta$ -D-glucopyranoside. Energy is given as a function of the distance between the ring carbon and the oxygen (marked C1 and O1 in the figure). (b) Energy barrier for the opening of the ether linkage of levoglucosan (1,6-anhydro- $\beta$ -D-glucopyranose). Energy is given as a function of the distance between C1 and O6 as marked in the figure

Figure 9: Dissociation of ether bond after protonation

the bond distance at the minimum. ReaxFF seems to give a longer bond length for the ether bond when the ether oxygen is protonated than what the DFT calculations show (about  $0.2 \text{ \AA}$  longer). We are mostly interested in reaction energies and thus we allow some deviations in the geometric structure, but this deviation is quite large and should be corrected for if possible in later parametrisations.

To conclude the protonation calculations, some stabilities of relevant charged glucose structures are given in Figure 10. Both the DFT calculations and ReaxFF show that the most stable protonated glucose is the structure where the ring oxygen is protonated (the one used as a reference structure in the figure). For the rest of the protonated structures, there is some disagreement between ReaxFF and DFT, especially in that ReaxFF does not favour the protonated O6-structure compared to the other protonated structures, as DFT does.

The carbocations modelled mostly turned into some odd structures, where only the carbocation at carbon 1 did not rearrange. ReaxFF seems to get the energy of the structure that rearranged to a furanose ring right, while the energy of the two other structures are not properly described. The carbocations not displayed in Figure 10 were not stable and did dissociate during the DFT geometry optimisation, and were therefore not included in the training set.

### 6.1.5 Carbohydrate conformations

Different conformers of carbohydrates have also been studied. The barrier for rotation around the  $C5 - C6$  bond of glucose has been studied and plotted in Figure 11(a). ReaxFF predicts this rotational barrier to be lower than the DFT-barrier for  $75^\circ \leq \theta \leq 150^\circ$ . This is believed to be an artifact of a too large interaction between a hydrogen on carbon 6 and the hydroxy hydrogen of carbon 4. The rotation barrier for rotation around the  $C1 - O1$  bond for methyl- $\beta$ -D-glucopyranoside has been calculated and plotted in Figure 11(b). Also the ReaxFF energies for this rotational barrier fits well, except for the single point where  $\theta = 0^\circ$ . Here the two hydrogens of the methyl carbon and the hydrogen of the hydroxy group of C6 are all about  $2.5 \text{ \AA}$  away from the ring-oxygen, which could give some level of stabilisation by hydrogen bonds. However, all in all the rotational barriers that ReaxFF predicts fit well to the DFT data both in the prediction of the

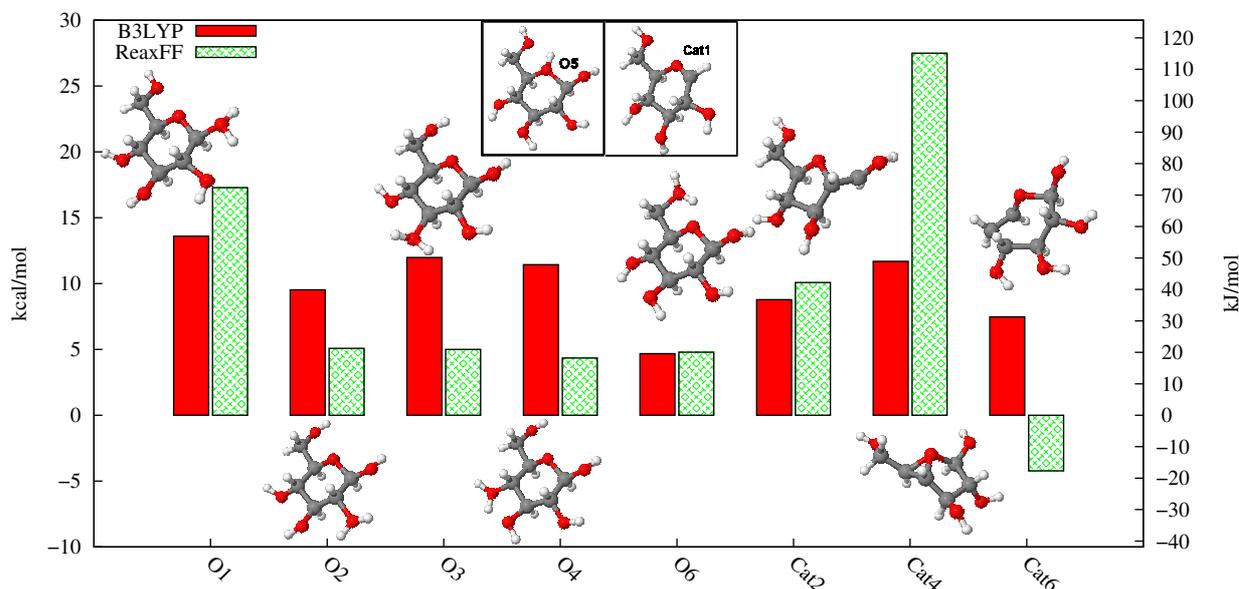
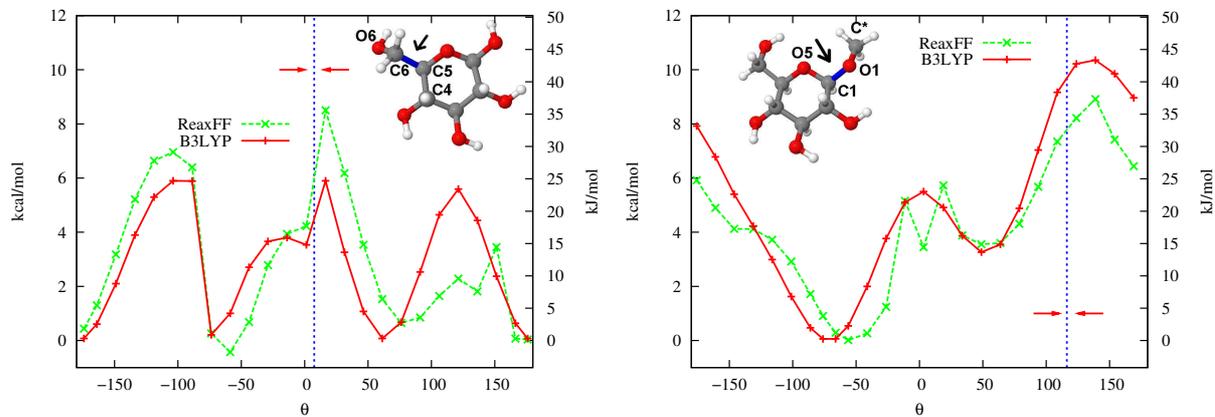


Figure 10: Energies of some charged glucose structures. Indices OX indicate protonation on oxygen X. Indices CatX indicate that  $\text{OH}^-$  has been removed from carbon X in the glucose ring, forming a carbocation. A glucose with the ether oxygen protonated is used as a reference for the protonated structures (O5), and a carbocation formed by removing a  $\text{OH}^-$  from carbon 1 is used as a reference for the carbocations (Cat1).

position of the minimas and the height of the energy barriers.

A transition from the  ${}^4\text{C}_1$  chair conformation to a  ${}^1\text{C}_4$  chair has been forced by restricting two dihedral angles in the ring sequentially. The energies as functions of those angles are shown in Figures 12(a) and 12(b). ReaxFF predicts the energy barrier of this rotation very well, the energies of the structures on the reaction path are quite similar, and also the energy of the final products of the transition is well described. The discrepancies in Figure 12(b) are a result of a rotation of a hydroxy group in the ReaxFF calculations which happened earlier than for the DFT calculations. The paths join again towards the end of the rotation.

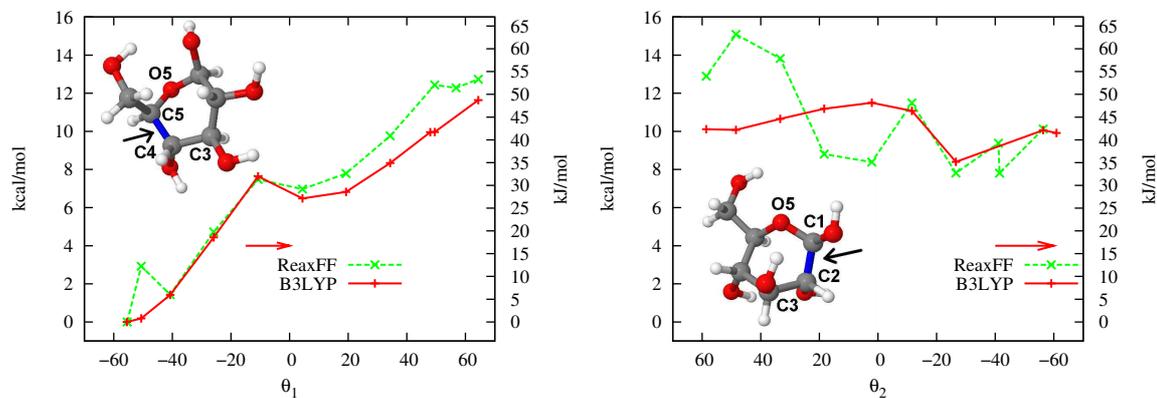
Finally, the stability of some conformers as reported by Appell et al.[67] has been compared, together with some isomers of glucofuranose and an open chain form of glucose in Figure 13. The energy differences between the different conformations of glucose are small, and ReaxFF is not able to get these small differences right. The well-known anomeric effect[95] which stabilises the  $\alpha$ -pyranoses is clearly shown by the DFT-calculations, but is not reproduced by ReaxFF. However ReaxFF correctly get that the  ${}^1\text{C}_4$  chair conformation is higher in energy than the  ${}^4\text{C}_1$  chair conformations. Also the furanoses are predicted to have higher energy than the pyranoses, though ReaxFF overpredicts this energy difference. ReaxFF is also able to get the energy difference between the pyranose form and the open form of glucose right.



(a) Energy barrier of rotation around the  $C5 - C6$  bond of  $\beta$ -D-glucopyranose. Angle  $\theta$  is defined as the dihedral angle  $C4 - C5 - C6 - O6$  as indicated in the figure, and the opposite pointing arrows indicate the direction of the rotation.

(b) Energy barrier of rotation around the  $C1 - O1$  bond of methyl- $\beta$ -D-glucopyranoside. Angle  $\theta$  is defined as the dihedral angle  $O5 - C1 - O1 - C^*$  as indicated in the figure, and the opposite pointing arrows indicate the direction of the rotation.

Figure 11: Rotational barriers for  $\beta$ -D-glucose and methyl- $\beta$ -D-glucopyranoside



(a) Energy as a function of the dihedral angle  $C3 - C4 - C5 - O5$  as shown in the figure. The arrow indicates the direction of rotation, and the rotation is started from the minimum structure of  $\beta$ -D-glucose.

(b) Energy as a function of the dihedral angle  $C3 - C2 - C1 - O5$  as shown in the figure. The arrow indicates the direction of rotation, and the rotation is started from a minima close to the final point of the rotation as performed in Figure 12(a). The reference energy is the energy of the minimised structure of  $\beta$ -D-glucose.

Figure 12: Energy of transition from  ${}^4C_1$  to  ${}^1C_4$  chair conformation of  $\beta$ -D-glucose

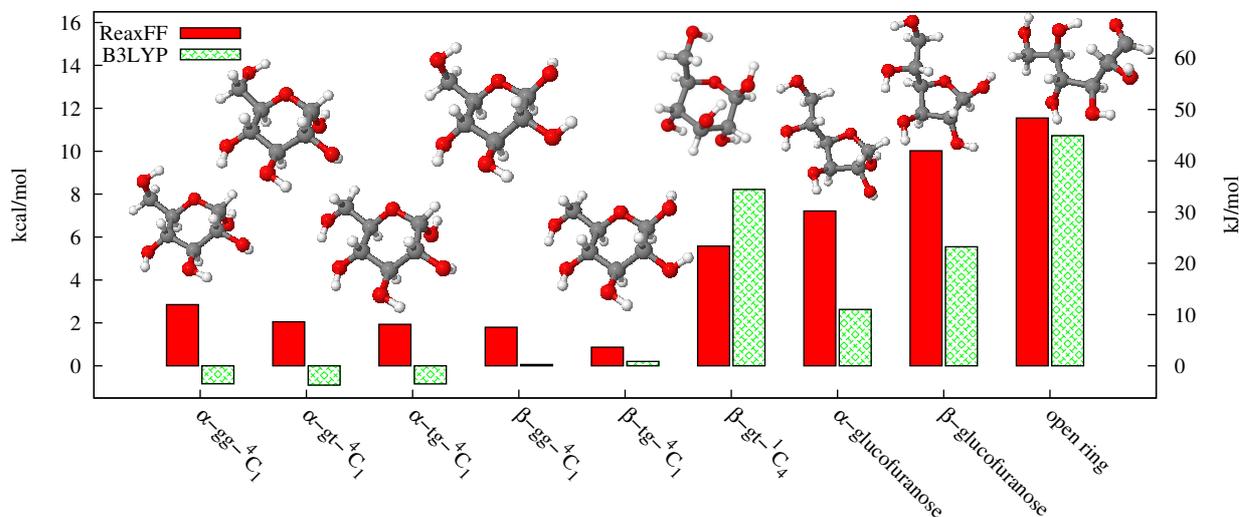


Figure 13: Energies of different isomers of glucose. The  ${}^4C_1$ -structures are identical to the structures reported by Appell et al.[67]. The  $\beta$ -gt- ${}^4C_1$  structure (Figure 4) has been used as reference structure, and the  $\beta$ -gt- ${}^1C_4$  structure is a  ${}^1C_4$ -chair version of the reference structure.

## 6.2 Test of force field

To demonstrate how a working force field can be employed in studying reactions and to probe the quality of the force field, a couple of relevant examples of its use have been made. These examples have also been studied in detail with the aim to find discrepancies in the force field, and parameters such as reaction products of simulations and reaction mechanisms have been investigated to check if they are reasonable.

### 6.2.1 Density of cellulose

The first test of the force field is to see if it can reproduce the density of cellulose. The crystal structure of cellulose denoted  $I_\alpha$  has been taken from diffraction data [93], and the predicted structure from ReaxFF has been aligned with and compared to diffraction data in Figure 14. The fit is mostly good, only two things are to be noted. ReaxFF predicts the ring angles to be a bit different, resulting in that the carbon 3 and its bonded oxygen of the monomer is pointing a bit more out of the plane than diffraction data gives. An energy plot for expansion/compression in the three unit cell directions (unit cell as defined in[93]) is shown in Figure 15. Compression and expansion in the X-direction directly compresses bonds, and thus give a high resistance to compression/expansion. This also seems to be the case for compression in the Z-direction. The expansion in the Y and Z directions give relatively flat curves, as this corresponds to increasing the distance between chains, and does not conflict with any bonds. Kinks on the curves correspond to small geometrical rearrangements, e.g. rotation of an O-H group.

The volume ReaxFF finds as a minimum for the unit cell lies about 10 % higher than the diffraction data gives. It should be noted that since the compression curve in the Y-direction is very flat (a reduction of the volume predicted by ReaxFF by 10 % only corresponds to an increase in energy of 3 kcal/mol), a small error in the energy description will lead to a relatively large error in the predicted volume. Thus the 10 % deviation is not worrying and ReaxFF performs satisfactory in the prediction of the unit cell.

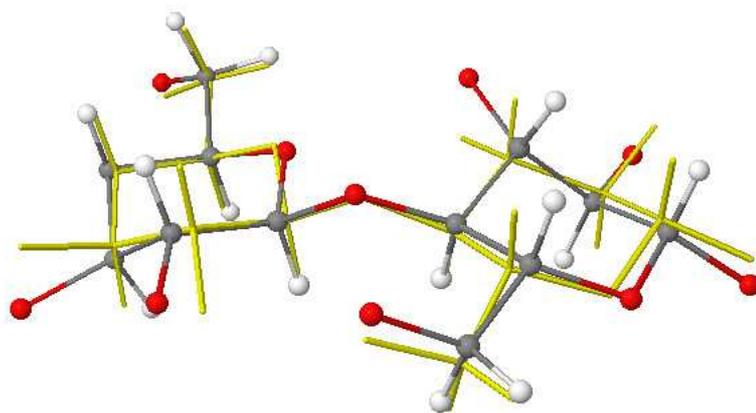


Figure 14: Crystal structure of cellulose  $I_\alpha$  optimised with ReaxFF (multicoloured and with atoms as balls, compared to diffraction crystal structure[93] (unicoloured)

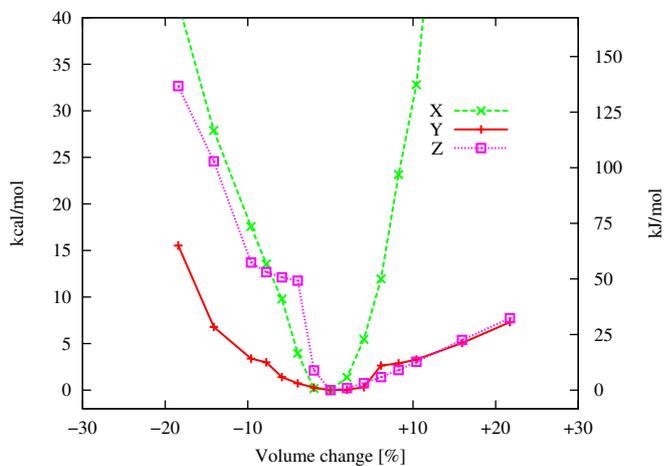
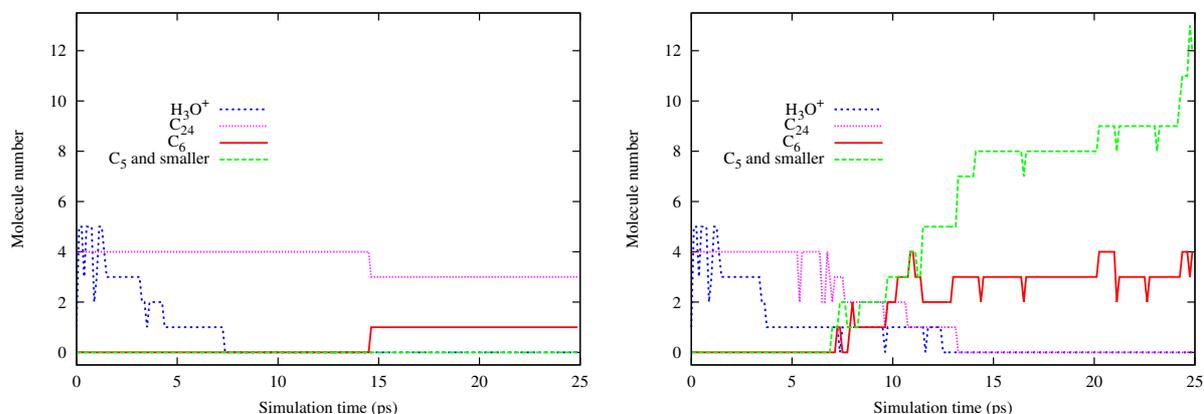


Figure 15: Energy change when unit cell length is decreased/increased in each of the three unit cell directions X, Y and Z. X is directed along the chain while both Y and Z is give directions between neighbouring chains. Refer to [93] for full unit cell definitions.

## 6.2.2 Simulations of the degradation of cellulose

The second example of what a full working force field is able to, is to simulate the degradation of cellulose. Such a simulation also gives us an opportunity to investigate the number of reactions that are predicted, and hopefully lets us identify elements to include for further parametrisations. The simulated systems consist of cellulose chains with a surrounding water phase including HCl (refer to Chapter 5 for details). It has to be noted that since only one simulation has been performed for each condition, starting from almost identical initial boxes, the presented results are probably highly correlated and do not show any correct statistics. The composition as a function of time for two such simulations is shown in Figure 16 for temperatures of 1000 K and 1500 K. The high temperature is a necessity to induce reactions on such a short time scale as the molecular dynamics simulations ( $< 1$  ns).



(a) Composition as a function of time for cellulose in HCl solution, starting with 5  $\text{H}_3\text{O}^+$  at 1000 K (b) Composition as a function of time for cellulose in HCl solution, starting with 5  $\text{H}_3\text{O}^+$  at 1500 K

Figure 16: Composition of simulated boxes containing cellulose strains surrounded by a water phase with varying concentration of HCl, cellulose/HCl(aq)

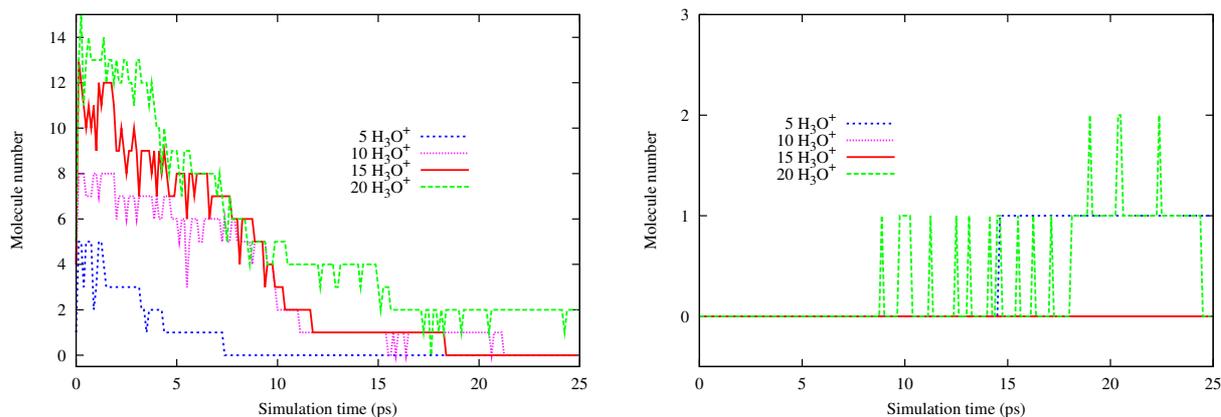


Figure 17: Presence of  $\text{H}_3\text{O}^+$  molecules during simulations of cellulose/HCl(aq) at 1000 K

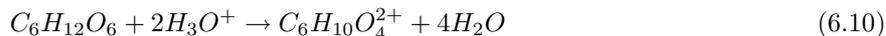
Figure 18: Presence of  $\text{C}_6$  fragments during simulations of cellulose/HCl(aq) at 1000 K

By looking at Figure 16 we see that at 1500 K, the degradation occurred faster and also produced smaller fragments than the degradation at 1000 K. An important variable seems to be the production of small carbon fragments, which are almost non-existent at 1000 K. The temperature dependence of those reactions indicates that they have a high energy barrier and need a higher temperature to take place. We are usually interested in describing reactions happening at a lower temperature (we only use the high temperature as a tool to speed up reactions), and if possible, eliminate such high energy reactions as the defragmentation of the  $C_6$  units. We will therefore only look at the reaction at 1000 K.

Another important observation from Figure 16 is that the  $H_3O^+$  in the solution is quickly consumed by the cellulose. This is expected as ReaxFF predicts that the transition of  $H^+$  from  $H_3O^+$  to cellulose is energetically favourable (as shown in Chapter 6.1.2). Simulations have been performed at different  $H_3O^+$  concentrations starting with 5, 10, 15 and 20  $H_3O^+$  ions (corresponding to a pH of about -0.3, -0.6, -0.8 and -0.9 respectively), and the consumption of  $H_3O^+$  during the simulations is shown in Figure 17. It seems that when starting with 20  $H_3O^+$  molecules we approach some kind of equilibrium, as the number of  $H_3O^+$  molecules is more or less constant over time. However, it should be noted that the simulation time is very short, and the system has probably not equilibrated at this timescale.

We now look at the production of  $C_6$  fragments as the indicator for the decomposition of cellulose. The production of  $C_6$  fragments for simulations starting with different amounts of  $H_3O^+$  is shown in Figure 18. This does not show any increased defragmentation by a larger initial  $H_3O^+$  concentration since defragmentation occurs in the simulations when starting with 5  $H_3O^+$  molecules and not in the simulations starting with 10 or 15  $H_3O^+$  molecules. This could signify that the defragmentation of cellulose is independent of the  $H_3O^+$  concentration, or basically that the protonation occurred at the wrong places in the cellulose chain, (remember that all the  $H_3O^+$  molecules were quickly consumed in these simulations). However, the presence of  $H_3O^+$  is important for the defragmentation, as simulations of only water and cellulose did not give any reactions at temperatures of 1000, 1250 and 1500 K.

By only looking at the end products of the simulations, one could argue that the force field predicts reasonable results. It is expected that the cellulose chains protonate and dissolve into  $C_6$  fragments. However we see that the protonation occurs very quickly, and that the cellulose consumes the  $H^+$  quickly. This is probably an artifact of the protonation reaction already treated in Chapter 6.1.2. However, by looking into the simulations in detail, a reaction of the type



occurred. By looking of the different steps, this reaction is an unphysical artifact of the force field, and this reaction will be addressed in the following chapter, but since ReaxFF predicts the products of this reaction to be stable, this explains why  $H_3O^+$  seems to disappear from the simulated systems (Figure 17).

### 6.2.3 Analyzing reaction paths

The newly implemented NEB-method (Chapter 4.4) enables us to easily obtain complicated reaction paths by using the ReaxFF force field, and also approximate the transition states. During the simulations as described in Chapter 6.2.2, an odd reaction (equation 6.10) occurred. A reaction of this type has been studied closer with the NEB-method, to demonstrate the use of the NEB-method and to point out a failure of the present force field.

It was observed that the following reaction occurred during the simulations

1. Protonation of glucose monomer at carbon 2
2. Dissociation of water, creating a carbocation at position 2
3. Protonation of glucose monomer at carbon 3
4. Dissociation of water, creating vicinal carbocations
5. Formation of a double bond from the vicinal carbocations

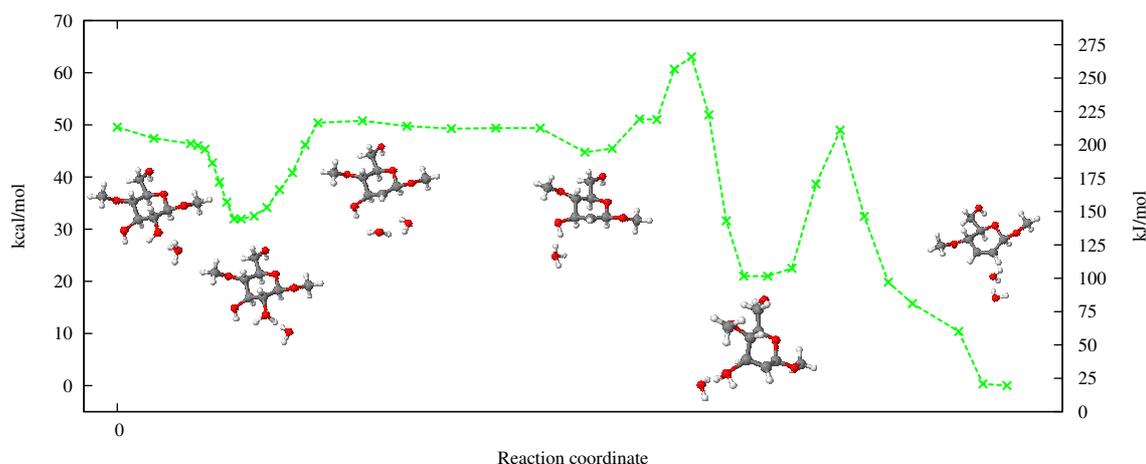


Figure 19: Reaction profile for the protonation and dissociation of 2 OH-groups on a methylated glucose unit.

These reaction steps for a system consisting of a glucose monomer, methylated at O1 and O4, together with a couple of surrounding water molecules, 2  $\text{H}_3\text{O}^+$  and 2  $\text{Cl}^-$  ions have been modelled by the NEB method, giving the reaction profile shown in Figure 19. This reaction profile points out some of the present problems of the ReaxFF parametrisation.

1. The first carbocation is stable  
The DFT calculations did not show any stable glucose carbocations, with the exception of the carbocation at carbon 1 which is stabilised by resonance with the ring oxygen. ReaxFF however creates these carbocations, and also protonates and dissociates a nearby hydroxy group, and creates two neighbouring carbocations. This reaction should be very unlikely on the limit to impossible, due to the two charges so close to each other. By looking at Figure 19 the first three structures give the formation of a carbocation. The carbocation is predicted by ReaxFF to be as stable as the original methylated glucose. The only carbocation structures included in the parametrisation were the relaxed intermediates as shown in Figure 10, since DFT did not give any other stable carbocation structures, thus a penalty for the structures ReaxFF predicts as stable should be introduced into the training set for further parametrisation.
2. A second carbocation is formed  
The second protonation shows a energy barrier of 15 kcal/mol, which is not impossible to overcome, but at least larger than the barrier of the first protonation. However, after the protonation, ReaxFF predicts a stable intermediate, 30 kcal/mol more stable than the initial methylated glucose. This is believed to be an artifact of that ReaxFF attempts to form a double bond between the carbocations.

### 3. Formation of a double bond

The second protonated hydroxy group is not easy to pull off and experiences an energy barrier of about 30 kcal/mol, and the resulting structure is a ring with a double bond. This end structure is predicted to be 50 kcal/mol more stable than the initial structure, thus giving the total reaction an exothermic character. The final glucose structure should in principle have a charge of +2, but the ReaxFF charges only show a total charge of +0.151 on the glucose-structure. Thus ReaxFF pushes the charge onto nearby water molecules in order to form a neutral glucose-structure. This final reaction is not physical, since the double bond is formed from electrons that are already removed by the dissociation of water.

The main problem seems to lie in the stability of the carbocation. If ReaxFF would give an unstable carbocation, this would dissociate before coming to the situation where a second carbocation and later a double bond is formed, thus removing the problem. The carbocations that ReaxFF predicts should therefore be given a penalty energy in further parametrisations, and this penalty energy can be estimated by calculating the DFT energy for a similar structure, keeping some coordinates constrained to prevent rearrangement.

## 6.3 Where to go next?

The present force field models energy barriers of oxidation and protonation well, and also gives good trends when it comes to the stability of some relevant structures of carbohydrates. However, it fails in two main ways which are quite important. A reparametrisation should therefore be performed to attempt to include these effects.

### 1. Stabilisation effect of water on cations

As shown in Chapter 6.1.2, the effect that water has on the protonation reaction is not modelled. This seems to be a problem with the stability of the  $\text{H}_3\text{O}^+$  molecule. The effect that neighbouring water has on the stability of  $\text{H}_3\text{O}^+$  should, if possible, be included in a new parametrisation. This could be performed including the DFT-calculations shown in Figure 7(b) in the parametrisation and by performing the energy comparison of  $\text{H}_2\text{O}/\text{H}_3\text{O}^+$  clusters as shown in equation 6.7.

### 2. Stability of carbocations

ReaxFF predicts carbocations to be stable intermediates. DFT-calculations on carbocation structures did in most cases not give stable intermediates, and did not give the same carbocation structures as ReaxFF. The carbocation structures that ReaxFF shows, should therefore be given a high energy in a reparametrisation. The only way ReaxFF distinguishes between radicals and cations, is by the overall charge of the system, or for a neutral system, the overall charge of parts of the system. If the EEM does not assign charges properly, a carbocation could react and behave like a radical, an event that was really observed in the simulations. This should be prevented, either by looking into the EEM parameters, or hoping that the reduced stability of carbocations reduce the probability that two carbocations interact.

These are the two main issues to solve in a further parametrisation of the force field. We allow certain inaccuracy in the force field, but these effects are quite crucial to most applications of such a force field. Another deviation from DFT calculations is the bond length of the protonated hydroxy groups on the sugar units. As described in Chapter 6.1.2 these are significantly longer than the DFT calculations show. This bond distance could indeed be longer than the present DFT calculations show, since Figure 7(b) shows that the bond length of the proton changes when water molecules are present. Nevertheless this should be an effect of nearby water molecules and not the bond itself.

### 6.3.1 Improvement of algorithms

The implemented Quasi-Newton method constructs a complete Hessian matrix by the BFGS updating procedure. For large systems, this matrix will grow large (scales as  $\mathcal{O}(N_{at}^2)$ ), and this could lead to memory issues.

An obvious improvement would be to only store half the matrix as it is symmetric, but a better approach will be to use the method of Nocedal[96, 97] where only the sets of gradients and coordinates from the last steps (typically about 5 steps) are stored and applied instead of the full Hessian matrix. This will reduce the memory requirement to a function of  $\mathcal{O}(N_{at})$  scaling. This will also be a useful improvement for the NEB-BFGS algorithm, where the memory requirement of the current implementation scales quadratically as a function of both the atom number and the number of images on the path ( $\mathcal{O}(N_{at}^2 N_{im}^2)$ ).

## 7 Conclusion

ReaxFF has been parametrised after a number of DFT calculations to reproduce protonation and oxidation barriers and to describe the stability of relevant carbohydrate structures. ReaxFF gives a good description of reaction barriers, shows good trends in the stability of glucose conformations and reproduce stabilities of many relevant carbohydrate structures. It estimates the density of cellulose I $_{\alpha}$  to be 10 % higher than diffraction data predicts, and simulations of cellulose in HCl solution at 1000 K predict depolymerisation of the cellulose chain into C $_6$  fragments. Two major defects of the present parametrisation are identified, the effect that H $_2$ O has on the stability of H $_3$ O $^+$  is not reproduced, and that ReaxFF gives too stable carbocations and lets them react as radicals, forming bonds between each other. It is believed that these effects can be prevented by a reparametrisation of the force field by inclusion of relevant structures.

The ReaxFF program package has been expanded by the implementation of a Quasi-Newton method for optimisation based on the BFGS-update, a Root Mean Square Distance fit procedure and a Nudged Elastic Band method for the location of reaction paths. Also a new formulation of the proof of how to project out translation and rotation from forces has been given.

## References

- [1] Lundgaard, L. E.; Hansen, W.; Linhjell, D.; Painter, T. J. *IEEE Trans. Power Deliv.* **2004**, *19*, 230-238.
- [2] Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerge, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719-1751.
- [3] Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.
- [4] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701-1718.
- [5] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorsid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Shulten, K. *J. Comput. Chem.* **2005**, *26*, 1781-1802.
- [6] van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard III, W. A. *J. Phys. Chem. A* **2001**, *105*, 9396-9409.
- [7] Ho, B. "The Root Mean Square Deviation", <http://www.dillgroup.ucsf.edu/~bosco/rmsd.html>, 2006 Downloaded 13.05.2006.
- [8] Kabsch, W. *Acta Cryst.* **1978**, *A34*, 827-828.
- [9] Mills, G.; Jónsson, H. *Phys. Rev. Lett.* **1994**, *72*, 1124-1127.
- [10] Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transition. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J.; Ciccotti, G.; Coker, D. F., Eds.; World Scientific: Singapore, 1998.
- [11] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833-1840.
- [12] El Khadem, H. S. *Carbohydrate Chemistry*; Academic Press Inc.: San Diego, CA, USA, 1<sup>st</sup> ed.; 1988.
- [13] Hounsell, E. *Chem. Int.* **2005**, *27*, 34.
- [14] Li, H.; Wang, L.-X. *Org. Biomol. Chem.* **2004**, *2*, 483-488.
- [15] Mohr, M.; Bryce, R. A.; Hiller, I. H. *J. Phys. Chem. A* **2001**, *105*, 8216-8222.
- [16] Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; van Gunsteren, W. F.; Hünenberger, P. H. *Biophys J.* **2006**, *90*, 4337-4344.
- [17] Dashnau, J. L.; Sharp, K. A.; Vanderkooi, J. M. *J. Phys. Chem. B* **2005**, *109*, 24152-24159.
- [18] Qian, X. *Carbohydr. Res.* **2005**, *340*, 2319-2327.
- [19] Abell, G. C. *Phys. Rev. B* **1985**, *31*, 6184-6196.
- [20] Tersoff, J. *Phys. Rev. B* **1988**, *37*, 6991-700.
- [21] Brenner, D. W. *Phys. Rev. B* **1990**, *42*, 9458-9471.
- [22] Stuart, S. J.; Tutein, A. B.; Harrison, J. A. *J. Chem. Phys.* **2000**, *112*, 6472-6486.

- [23] Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. *J. Phys.: Condens. Matter* **2002**, *14*, 783-802.
- [24] van Duin, A. C. T.; Strachan, A.; Stewman, S.; Zhang, Q.; Xu, X.; Goddard III, W. A. *J. Phys. Chem. A* **2003**, *107*, 3803-3811.
- [25] Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W. Q.; Goddard III, W. A. *J. Phys. Chem. A* **2005**, *109*, 493-499.
- [26] Strachan, A.; van Duin, A. C. T.; Chakraborty, D.; Dasgupta, S.; Goddard III, W. A. *Phys. Rev. Lett.* **2003**, *91*, 098301.
- [27] Strachan, A.; Kober, E. M.; van Duin, A. C. T.; Oxgaard, J.; Goddard III, W. A. *J. Chem. Phys.* **2005**, *122*, 054502.
- [28] van Duin, A. C. T.; Zeiri, Y.; Dubnikova, F.; Kosloff, R.; Goddard III, W. A. *J. Am. Chem. Soc.* **2005**, *127*, 11053-11062.
- [29] Chen, N.; van Duin, A. C. T.; Goddard III, W. A.; Lusk, M. T. *Phys. Rev. B* **2005**, *72*, 085416.
- [30] Pouwels, A. D.; Eijkel, G. B.; Boon, J. J. *J. Anal. Appl. Pyrol.* **1988**, *14*, 237-280.
- [31] McQuarrie, D. A.; Simon, J. D. *Physical Chemistry: A Molecular Approach*; Univ Science Books: Sausalito, CA, USA, 1<sup>st</sup> ed.; 1997.
- [32] Jensen, F. *Introduction to Computational Chemistry*; John Wiley and Sons: West Sussex, England, 1<sup>st</sup> ed.; 1999.
- [33] Capelle, K. "A bird's-eye view of density-functional theory", <http://arxiv.org/abs/cond-mat/0211443>, 2005.
- [34] Kohanoff, J.; Gidopoulos, N. I. Density Functional Theory: Basics, New Trends and Applications. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; John Wiley & Sons Ltd: Chichester, 2003.
- [35] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B 864-871.
- [36] Kato, T. *Commun. Pure Appl. Math.* **1957**, *10*, 151-177.
- [37] Feynman, R. P.; Leighton, R. B.; Sands, M. *The Feynman Lectures on Physics*; volume 2, chapter 19 Addison-Wesley Publishing Company: Reading, Massachusetts, 1964.
- [38] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133-A1138.
- [39] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [40] Becke, A. D. *Phys. Rev.* **1988**, *38*, A3098-3100.
- [41] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211 Also available from <http://www.cas.org/spotlight/bchem02/200307j6.html>.
- [42] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *B 37*, 785-789.
- [43] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
- [44] Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345-351.

- [45] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
- [46] Gell-Mann, M.; Brckner, K. A. *Phys. Rev.* **1957**, *106*, 364-368.
- [47] Frisch, M. J. *et al.* "Gaussian 03, Revision C.02", Gaussian, Inc., Wallingford, CT, 2004.
- [48] "Jaguar, Version 6.5", Schrodinger, LLC, New York, NY, 2005.
- [49] Colle, R.; Salvetti, O. *Theor. Chim. Acta* **1975**, *37*, 329-334.
- [50] Slater, J. C. *Phys. Rev.* **1930**, *36*, 57-64.
- [51] Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657-2664.
- [52] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650-654.
- [53] Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1984**, *80*, 3265-3269.
- [54] Spitznagel, G. W.; Clark, T.; Chandrasekhar, J.; v. R. Schleyer, P. *J. Comput. Chem.* **1981**, *3*, 363-371.
- [55] Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; v. R. Schleyer, P. *J. Comput. Chem.* **1983**, *4*, 294-301.
- [56] Bene, J. E. D.; Jordan, M. J. T. *J. Mol. Struct. (Theochem)* **2001**, *573*, 11-23.
- [57] Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553-566.
- [58] Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024-11031.
- [59] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063-1079.
- [60] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221-7230.
- [61] Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2001**, *114*, 5497-5503.
- [62] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.
- [63] Bernardi, F.; Bottoni, A.; Garavelli, M. *QSAR & Combinatorial Science* **2002**, *21*, 128-148.
- [64] Momany, F. A.; Willett, J. L. *J. Comput. Chem.* **2000**, *21*, 1204-1219.
- [65] Lii, J.-H.; Ma, B.; Allinger, N. L. *J. Comput. Chem.* **1999**, *20*, 1593-1603.
- [66] Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L. *Carbohydr. Res.* **2004**, *339*, 553-567.
- [67] Appell, M.; Strati, G.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2004**, *339*, 537-551.
- [68] Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955-1970.
- [69] Chenoweth, K.; Cheung, S.; van Duin, A. C. T.; Goddard III, W. A.; Kober, E. M. *J. Am. Chem. Soc.* **2005**, *127*, 7192-7202.
- [70] Cheung, S.; Deng, W. Q.; van Duin, A. C. T.; Goddard III, W. A. *J Phys. Chem. A* **2005**, *109*, 851-859.
- [71] Csonka, G. I. *J Mol Struct (Theochem)* **2002**, *584*, 1-4.
- [72] Salvador, P.; Fradera, X.; Duran, M. *J. Chem. Phys.* **2000**, *112*, 10106-10115.

- [73] van der Vaart, A.; Merz, Jr., K. M. *J. Chem. Phys.* **2002**, *116*, 7380-7388.
- [74] Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618-622.
- [75] Åstrand, P.-O.; Linse, P.; Karlström, G. *Chem. Phys.* **1995**, *191*, 195-202.
- [76] Mortier, W. J.; Ghosh, S. K.; Shankar, S. *J. Am. Chem. Soc.* **1986**, *108*, 4315-4320.
- [77] Rappé, A. K.; Goddard III, W. A. *J. Phys. Chem.* **1991**, *95*, 3358-3363.
- [78] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637-649.
- [79] Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.
- [80] Feynman, R. P.; Leighton, R. B.; Sands, M. *The Feynman Lectures on Physics*; volume 1, chapter 9 Addison-Wesley Publishing Company: Reading, Massachusetts, 1963.
- [81] Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082-2094.
- [82] Broyden, C. G. *Not. AMS* **1969**, *16*, 670.
- [83] Fletcher, R. *Comput. J.* **1970**, *13*, 317-322.
- [84] Goldfarb, D. *Math. Comp.* **1970**, *24*, 23-26.
- [85] Shanno, D. F. *Math. Comp.* **1970**, *24*, 647-657.
- [86] Powell, M. J. D. *Math. Prog.* **1971**, *1*, 26-57.
- [87] Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800-2806.
- [88] Banerjee, A.; Adams, N.; Shepard, R.; Simons, J. *J. Phys. Chem.* **1985**, *89*, 52-57.
- [89] Baker, J. *J. Comput. Chem* **1986**, *7*, 385-395.
- [90] Fletcher, R. *Practical Methods of Optimization*; Wiley: Chichester, UK, 2<sup>nd</sup> ed.; 2000.
- [91] Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978-9985.
- [92] Frisch, M. J. *et al.* "Gaussian 98, Revision A.9", Gaussian, Inc., Pittsburgh, PA, 1998.
- [93] Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300-14306.
- [94] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- [95] Juaristi, E.; Cuevas, G. *Tetrahedron* **1992**, *48*, 5019-5087.
- [96] Nocedal, J. *Math. Comput.* **1980**, *35*, 773-782.
- [97] Liu, D.; Nocedal, J. *Math. Program.* **1989**, *45*, 503-528.

## APPENDIX

### A ReaxFF Energy terms

The energy terms of ReaxFF are described in the supporting information to the article by Nielson et al.[25]. ReaxFF employs a large number of parameters, and a list of those parameters is given in Chapter A.5. The total energy of the ReaxFF is given as

$$E_{system} = E_{bond} + E_{lp} + E_{over} + E_{under} + E_{val} + E_{pen} + E_{coa} + E_{tors} + E_{conj} + E_{Hb} + E_{vdw} + E_{Coul} \quad (A.1)$$

where  $E_{bond}$ ,  $E_{lp}$ ,  $E_{over}$  and  $E_{under}$  are described in equations 3.21,3.24,3.27, and 3.28 respectively.

#### A.1 Valence angle energy

The valence angle energy for the angle  $\theta_{ijk}$  is given by the following equations

$$E_{val} = f_7(BO_{ij})f_7(BO_{jk})f_8(\Delta_j^{ang})[p_{v,1} - p_{v,1}e^{-p_{v,2}(\Theta_0(SBO_2) - \theta_{ijk})^2}] \quad (A.2)$$

$$f_7(BO_{ij}) = 1 - e^{-p_{v,3}BO_{ij}^{p_{v,4}}} \quad (A.3)$$

$$f_8(\Delta_j^{ang}) = p_{v,5} - (p_{v,5} - 1) \frac{2 + e^{p_{v,6}\Delta_j^{ang}}}{1 + e^{p_{v,6}\Delta_j^{ang}} + e^{-p_{v,7}\Delta_j^{ang}}} \quad (A.4)$$

$$SBO = \sum_{n=1}^{nb(j)} (BO_{jn}^{\pi} + BO_{jn}^{\pi\pi}) + (1 - \prod_{n=1}^{nn(j)} e^{-BO_{jn}^s})(-\Delta_j^{ang} - p_{v,8}nl_{p,j}) \quad (A.5)$$

$$\Delta_j^{ang} = \sum_{n=1}^{nb(j)} BO_{jn} - Val_j^{ang} \quad (A.6)$$

$$\begin{aligned} SBO_2 &= 0 && \text{if } SBO \leq 0 \\ SBO_2 &= SBO^{p_{v,9}} && \text{if } 0 < SBO \leq 1 \\ SBO_2 &= 2 - (2 - SBO)^{p_{v,9}} && \text{if } 1 < SBO < 2 \\ SBO_2 &= 2 && \text{if } 2 \leq SBO \end{aligned} \quad (A.7)$$

$$\Theta_0 = \pi - \Theta_{0,0}(1 - e^{-p_{v,10}(2 - SBO_2)}) \quad (A.8)$$

where  $p_{v,i}$  are some parameters and  $Val_{ang} = Val^{ov}$  for non-metals.  $SBO$  defines the amount of  $\pi$ -character of the bonds, thus defining the hybridisation of the central atom. This is necessary, as the reference angle  $\Theta_0$  is dependent on the hybridisation of the central atom (should be approximately 109.5 deg for a  $sp^3$  carbon, while 120 deg for a  $sp^2$  carbon). The reference angle is plotted against  $SBO_2$  for a C-C-C angle in Figure A.1, and we see that it varies from about 105° at  $SBO_2 = 0$  to 180° for  $SBO_2 = 2$ . The bond order  $BO$  is also included in the energy calculations, as the valence angle energy should be turned off when atoms are not bonded.

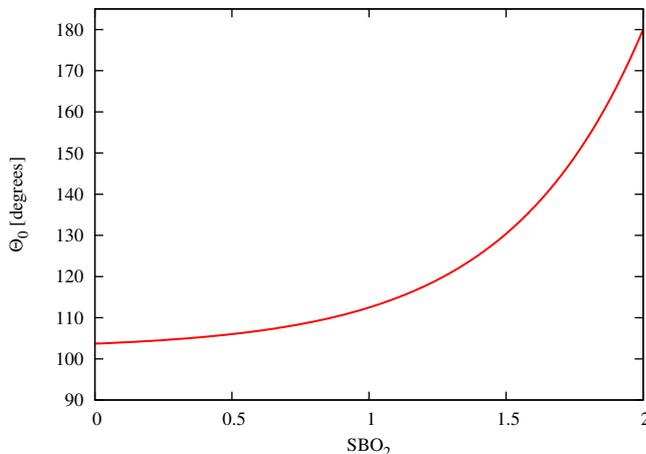


Figure A.1: The reference angle  $\Theta_0$  plotted against  $SBO_2$  for a C-C-C angle

## A.2 Torsion angle energy

The torsion angle energy for an angle  $\omega_{ijkl}$  is calculated by the following equations

$$E_{tors} = f_{10}(BO_{ij}, BO_{jk}, BO_{kl}) \sin \theta_{ijk} \sin \theta_{jkl} \left[ \frac{1}{2} V_2 e^{p_{t,1}(BO_{jk}^\pi - 1 + f_{11}(\Delta_j, \Delta_k))^2} (1 - \cos(2\omega_{ijkl})) + \frac{1}{2} V_3 (1 + 3 \cos(3\omega_{ijkl})) \right] \quad (\text{A.9})$$

$$f_{10}(BO_{ij}, BO_{jk}, BO_{kl}) = [1 - e^{-p_{t,2}BO_{ij}}][1 - e^{-p_{t,2}BO_{jk}}][1 - e^{-p_{t,2}BO_{kl}}] \quad (\text{A.10})$$

$$f_{11}(\Delta_j, \Delta_k) = \frac{2 + e^{-p_{t,3}(\Delta_j^{ang} + \Delta_k^{ang})}}{1 + e^{-p_{t,3}(\Delta_j^{ang} + \Delta_k^{ang})} + e^{p_{t,4}(\Delta_j^{ang} + \Delta_k^{ang})}} \quad (\text{A.11})$$

where the  $p_{t,i}$  are some parameters. The cosine functions give the rotation energy terms similar to non-reactive force fields (Chapter 3.1.2). The torsion angle energy is also scaled down when the bonds are breaking by the factor  $f_{10}$ . The  $f_{11}$  factor acts to increase the  $(1 - \cos(2\omega_{ijkl}))$ -term, thus increasing the rotational barriers for systems with high bond order, such as double bonds.

## A.3 Nonbonded interactions

The non-bonded interactions are actually simpler for ReaxFF than for conventional non-reactive force fields. Most non-reactive force fields don't include non-bonded interactions between bonded atoms, while ReaxFF treats all atoms equally. Thus the non-bonded interactions are calculated for all atom pairs.

A cutoff is used for the long range coulomb forces and van der Waals forces, and we define a taper function to make a continuous transition from the low magnitude energy at the cutoff to zero interaction energy. We do this by introducing a 7<sup>th</sup> order polynomial

$$T = 20 \frac{r^7}{R_{cut}^7} - 70 \frac{r^6}{R_{cut}^6} + 84 \frac{r^5}{R_{cut}^5} - 35 \frac{r^4}{R_{cut}^4} + 1 \quad (\text{A.12})$$

The coulomb energy is shielded at low distances, and is given as

$$E_{coul} = T \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{[r_{ij}^3 + (\frac{1}{\gamma_{ij}})^3]^{\frac{1}{3}}} \quad (\text{A.13})$$

where  $\gamma_{ij}$  is a shielding parameter. Notice that equation A.13 is given in SI units, thus the distance has to be given in  $m$  and the charges in  $C$ . The van der Waals energy is also shielded for short distances and is given as

$$E_{vdw} = T D_{ij} [e^{p_{vdw,1}(1 - \frac{f_{13}(r_{ij})}{r_{vdw}})} - 2e^{\frac{p_{vdw,1}}{2}(1 - \frac{f_{13}(r_{ij})}{r_{vdw}})}] \quad (\text{A.14})$$

where

$$f_{13}(r_{ij}) = [r_{ij}^{p_{vdw,2}} + (\frac{1}{\gamma_{vdw}})^{p_{vdw,2}}]^{\frac{1}{p_{vdw,2}}} \quad (\text{A.15})$$

where  $\gamma_{vdw}$  is a shielding parameter and  $p_{vdw,1}$  and  $p_{vdw,2}$  are parameters. The hydrogen bonds for a system are accounted for by the following energy term for a  $X - H \cdots Z$  - system.

$$E_{Hb} = p_{hb,1} [1 - e^{p_{hb,2} BO_{XH}}] e^{p_{hb,3} (\frac{r_{hb}^0}{r_{HZ}} + \frac{r_{HZ}}{r_{hb}^0} - 2)} \sin^8 \frac{\theta_{XHZ}}{2} \quad (\text{A.16})$$

for some parameters  $p_{hb,1}, p_{hb,2}, p_{hb,3}$  and  $r_{hb}^0$ .

#### A.4 Special terms

The three terms  $E_{pen}, E_{coa}$  and  $E_{conj}$  from equation A.1 have not yet been accounted for, and are corrections for special systems. Two terms have been introduced to stabilize conjugated systems. The first describes a I-J-K system and is defined as

$$E_{coa} = p_{coa,1} \frac{1}{1 + e^{p_{coa,2} \Delta_{ij}^{bo}}} e^{-p_{coa,3} (-BO_{ij} + \sum_{n=1}^{nb(i)} BO_{in})^2} e^{-p_{coa,3} (-BO_{jk} + \sum_{n=1}^{nb(k)} BO_{nk})^2} e^{-p_{coa,4} (BO_{ij} - 1.5)^2} e^{-p_{coa,4} (BO_{jk} - 1.5)^2} \quad (\text{A.17})$$

where  $p_{coa,i}$  are some parameters, and the second one describes a I-J-K-L system and is defined as

$$E_{conj} = f_{12}(BO_{ij}, BO_{jk}, BO_{kl}) p_{cot,1} [1 + (\cos^2 \omega_{ijkl} - 1) \sin \theta_{ijk} \theta_{jkl}] \quad (\text{A.18})$$

$$f_{12}(BO_{ij}, BO_{jk}, BO_{kl}) = e^{-p_{cot,2} (BO_{ij} - 1.5)^2} e^{-p_{cot,2} (BO_{jk} - 1.5)^2} e^{-p_{cot,2} (BO_{kl} - 1.5)^2} \quad (\text{A.19})$$

where  $p_{cot,i}$  are some parameters. Notice that both these energy terms have contributions of the form  $e^{-p(BO-1.5)^2}$ , which will have maximum for  $BO = 1.5$  which corresponds to a conjugated system. The final special term is an energy penalty for a system with two vicinal double bonds (I=J), and is defined as

$$E_{pen} = p_{pen,1} f_9(\Delta_j) e^{-p_{pen,2}(BO_{ij}-2)^2} e^{-p_{pen,2}(BO_{jk}-2)^2} \quad (\text{A.20})$$

$$f_9(\Delta_j) = \frac{2 + e^{-p_{pen,3}\Delta_j}}{1 + e^{-p_{pen,3}\Delta_j} + e^{-p_{pen,4}\Delta_j}} \quad (\text{A.21})$$

where  $p_{pen,i}$  are some parameters.

All energy contributions from equation A.1 are now described by the equations 3.21,3.24,3.27, 3.28,A.2, A.9, A.13, A.14, A.16, A.17, A.18 and A.20.

## A.5 ReaxFF parameters

ReaxFF employs a large number of parameters and functions which are gathered and listed below together with a small description and a reference to their first occurrence.

### List of ReaxFF Parameters

	Property	Unit	1 <sup>st</sup> Occurrence		
			Equation	Page	
$nb(i)$	Neighbours of atom $i$	none	3.8	18	
$int(x)$	Returns $x$ without decimals	none	3.22	21	
Derivated parameters					
$\Delta_i$	Overcoordination of atom $i$	none	3.13	19	
$\Delta_i^e$	Sum of bond order - number of valence electrons	none	3.22	21	
$\Delta_i^{lp}$	Deficiency of lone pairs	none	3.25	22	
$\Delta_i^{lpc}$	Overcoordination corrected for $\Delta_i^{lp}$	none	3.26	22	
$\Delta_i^{ov}$	Overcoordination including lone pairs	none	3.14	19	
$\Theta_0(SBO_2)$	Reference valence angle	rad	A.2	53	
$BO_{ij}$	Total bond order for bond $i - j$	$0 \leq BO_{ij} \leq 3$	none	3.12	18
$BO_{ij}^\sigma$	$\sigma$ contribution to $BO_{ij}$	$0 \leq BO_{ij}^\sigma \leq 1$	none	3.12	18
$BO_{ij}^\pi$	1 <sup>st</sup> $\pi$ contribution to $BO_{ij}$	$0 \leq BO_{ij}^\pi \leq 1$	none	3.12	18
$BO_{ij}^{\pi\pi}$	2 <sup>nd</sup> $\pi$ contribution to $BO_{ij}$	$0 \leq BO_{ij}^{\pi\pi} \leq 1$	none	3.12	18
$BO_{ij}^{\sigma,corr}$	Corrected $BO_{ij}^\sigma$	$0 \leq BO_{ij}^{\sigma,corr} \leq 1$	none	3.15	19
$BO_{ij}^{\pi,corr}$	Corrected $BO_{ij}^\pi$	$0 \leq BO_{ij}^{\pi,corr} \leq 1$	none	3.15	19
$BO_{ij}^{\pi\pi,corr}$	Corrected $BO_{ij}^{\pi\pi}$	$0 \leq BO_{ij}^{\pi\pi,corr} \leq 1$	none	3.15	19
$N_i^{eval}$	Number of valence electrons of atom $i$	none	3.23	21	
$n_{lp}$	Calculated number of lone pairs	none	3.22	21	
$n_{lp,i}^0$	Ideal number of lone pairs	none	3.25	22	
$R_{cut}$	Cut-off radius for long-range energy	none	A.12	55	
$SBO$	Degree of $\pi$ -bond	$0 \leq SBO \leq 2$	none	A.7	53
$SBO_2$	Function to calculate SBO	none	A.7	53	

$T$	Taper function, makes cut-off continuous.	none	A.12	55
$Val_i$	Valence of atom $i$ . Equals the number of bonds it forms	none	3.13	19
$Val_j^{ang}$	$Val_i$ for non-metals. Parameter for metals	none	A.5	53
$Val_i^{ov}$	$Val_i$ + ideal number of lone pairs	none	3.14	19
$\Delta_j^{ang}$	Overcoordination for angles	none	A.4	53

### Fitted parameters

$\gamma_{ij}$	Shielding parameter for coulomb energy		$\frac{1}{length}$	A.13	55
$\gamma_{vdw}$	Van der Waal's shielding parameter		$\frac{1}{length}$	A.15	55
$D_{ij}$	Van der Waal's dissociation parameter		$\frac{kcal}{mol}$	A.14	55
$D_e^\sigma$	Dissociation energy $\sigma$ bond	$D_e^\sigma \geq 0$	$\frac{kcal}{mol}$	3.21	21
$D_e^\pi$	Dissociation energy 1 <sup>st</sup> $\pi$ bond	$D_e^\pi \geq 0$	$\frac{kcal}{mol}$	3.21	21
$D_e^{\pi\pi}$	Dissociation energy 2 <sup>nd</sup> $\pi$ bond	$D_e^{\pi\pi} \geq 0$	$\frac{kcal}{mol}$	3.21	21
$p_{bo,1}$	Parameter 1 - Bond order	$p_{bo,1} < 0$	none	3.12	18
$p_{bo,2}$	Parameter 2 - Bond order	$p_{bo,2} > 0$	none	3.12	18
$p_{bo,3}$	Parameter 3 - Bond order	$p_{bo,3} < 0$	none	3.12	18
$p_{bo,4}$	Parameter 4 - Bond order	$p_{bo,4} > 0$	none	3.12	18
$p_{bo,5}$	Parameter 5 - Bond order	$p_{bo,5} < 0$	none	3.12	18
$p_{bo,6}$	Parameter 6 - Bond order	$p_{bo,6} > 0$	none	3.12	18
$p_{be,1}$	Parameter 1 - Bond energy	$p_{be,1} > 0$	none	3.21	21
$p_{be,2}$	Parameter 2 - Bond energy	$p_{be,2} > 0$	none	3.21	21
$p_{boc,1}$	Parameter 1 $BO$ -correction	$p_{boc1} > 0$	none	3.17	20
$p_{boc,2}$	Parameter 2 $BO$ -correction	$p_{boc2} > 0$	none	3.18	20
$p_{boc,3}$	Parameter 3 $BO$ -correction	$p_{boc3} > 0$	none	3.19	20
$p_{boc,4}$	Parameter 4 $BO$ -correction	$p_{boc4} > 0$	none	3.19	20
$p_{boc,5}$	Parameter 5 $BO$ -correction	$p_{boc5} > 0$	none	3.19	20
$p_{coa,1}$	Parameter 1 Conjugation A		$\frac{kcal}{mol}$	A.17	55
$p_{coa,2}$	Parameter 2 Conjugation A		none	A.17	55
$p_{coa,3}$	Parameter 3 Conjugation A		none	A.17	55
$p_{coa,4}$	Parameter 4 Conjugation A		none	A.17	55
$p_{cot,1}$	Parameter 1 Conjugation B		$\frac{kcal}{mol}$	A.18	55
$p_{cot,2}$	Parameter 2 Conjugation B		none	A.19	55
$p_{hb,1}$	Parameter 1 Hydrogen bond		$\frac{kcal}{mol}$	A.16	55
$p_{hb,2}$	Parameter 2 Hydrogen bond		none	A.16	55
$p_{hb,3}$	Parameter 3 Hydrogen bond		none	A.16	55
$p_{lp,1}$	Parameter to make $n_{lp,i}$ continuous		none	3.22	21
$p_{lp,2}$	Parameter 2 for lone pairs	$p_{lp,2} > 0$	$\frac{kcal}{mol}$	3.24	21
$p_{ov,1}$	Parameter 1 for overcoordination	$p_{ov,1} > 0$	none	3.27	22
$p_{ov,2}$	Parameter 2 for overcoordination	$p_{ov,2} < 0$	none	3.27	22
$p_{ov,3}$	Parameter 3 for overcoordination	$p_{ov,3} > 0$	none	3.26	22
$p_{ov,4}$	Parameter 4 for overcoordination	$p_{ov,4} > 0$	none	3.26	22
$p_{pen,1}$	Parameter 1 Penalty energy		$\frac{kcal}{mol}$	A.20	56
$p_{pen,2}$	Parameter 2 Penalty energy		none	A.20	56
$p_{pen,3}$	Parameter 3 Penalty energy		none	A.21	56
$p_{pen,4}$	Parameter 4 Penalty energy		none	A.21	56
$p_{t,1}$	Parameter 1 - Torsion angle energy		none	A.9	54
$p_{t,2}$	Parameter 2 - Torsion angle energy		none	A.10	54

$p_{t,3}$	Parameter 3 - Torsion angle energy		none	A.10	54
$p_{t,4}$	Parameter 4 - Torsion angle energy		none	A.10	54
$p_{un,1}$	Parameter 1 for undercoordination	$p_{un,1} > 0$	$\frac{kcal}{mol}$	3.28	22
$p_{un,2}$	Parameter 2 for undercoordination	$p_{un,2} < 0$	none	3.28	22
$p_{un,3}$	Parameter 3 for undercoordination	$p_{un,3} > 0$	none	3.28	22
$p_{un,4}$	Parameter 4 for undercoordination	$p_{un,4} < 0$	none	3.28	22
$p_{v,1}$	Parameter 1 - Valence angle energy		$\frac{kcal}{mol}$	A.2	53
$p_{v,2}$	Parameter 2 - Valence angle energy		none	A.2	53
$p_{v,3}$	Parameter 3 - Valence angle energy		none	A.3	53
$p_{v,4}$	Parameter 4 - Valence angle energy		none	A.3	53
$p_{v,5}$	Parameter 5 - Valence angle energy		none	A.4	53
$p_{v,6}$	Parameter 6 - Valence angle energy		none	A.4	53
$p_{v,7}$	Parameter 7 - Valence angle energy		none	A.4	53
$p_{v,8}$	Parameter 8 - Valence angle energy		none	A.5	53
$p_{v,9}$	Parameter 9 - Valence angle energy		none	A.7	53
$p_{v,10}$	Parameter 10 - Valence angle energy		none	A.8	53
$p_{vdw,1}$	Parameter 1 Van der Waal's energy		none	A.14	55
$p_{vdw,2}$	Parameter 2 Van der Waal's energy		none	A.15	55
$r_{\sigma}$	Reference $\sigma$ bond distance		length	3.12	18
$r_{\pi}$	Reference 1 <sup>st</sup> $\pi$ bond distance		length	3.12	18
$r_{\pi\pi}$	Reference 2 <sup>nd</sup> $\pi$ bond distance		length	3.12	18
$r_{hb}^0$	Reference distance Hydrogen bond		none	A.16	55
$r_{vdw}$	Van der Waal's radius		length	A.14	55
$V_2$	Energy barrier parameter - Torsion angle energy		none	A.9	54
$V_3$	Energy barrier parameter - Torsion angle energy		none	A.9	54